人工智能的可持续融合：赋能教育、商业与健康的路径与治理
THE SUSTAINABLE INTEGRATION OF ARTIFICIAL INTELLIGENCE: PATHWAYS AND GOVERNANCE FOR EMPOWERING EDUCATION, BUSINESS, AND WELLNESS

崔丹 [1*], 袁颖姣 [2]
Dan Cui[1*] and Yingjiao Yuan[2]

[1,2] 中国哈尔滨工程大学国际合作教育学院
[1,2]College of International Cooperation and Education, Harbin Engineering University, China
[*]**Corresponding Author, E-mail:** cuidan@hrbeu.edu.cn

摘要

人工智能深度融入教育、商业与健康领域，是构建可持续智能未来的关键路径。现有研究多局限于单一领域，缺乏跨领域的系统性比较与整合性治理框架。本研究采用系统性文献综述方法，首次将三大领域置于“技术赋能——风险规制——可持续治理”的统一分析框架下，解析 AI 赋能模式、核心价值与共性挑战。研究发现，AI 在三大领域均展现出显著的赋能潜力，但也面临数据隐私、算法偏见、数字鸿沟等共性风险，其根源在于“技术赋能”与“自身可持续性”之间的内在张力。本研究的核心贡献在于：构建了“可持续智能未来”的三维度评估框架；揭示了 AI 作为“矛盾复合体”的辩证逻辑；提出了“共性框架+差异化策略”的双层治理体系。研究据此指出，实现可持续智能未来必须超越效率至上的工具理性，构建涵盖敏捷治理、伦理设计与包容性发展的系统性框架。

关键词： 人工智能 可持续发展 教育 商业 健康

Abstract

The deep integration of Artificial Intelligence (AI) into the fields of education, business, and wellness constitutes a pivotal pathway toward building a sustainable smart future. Existing research is largely confined to individual domains, lacking systematic cross-domain comparisons and an integrated governance framework. This study employs a systematic literature review methodology to, for the first time, place these three major fields within a unified analytical framework of “technology empowerment—risk regulation—sustainable governance” and analyzes AI empowerment models, core values, and common challenges. The findings reveal that while AI demonstrates significant empowerment potential across the three sectors, it also faces common risks such as data privacy concerns, algorithmic bias, and the digital divide. These risks stem from the inherent tension between “technological empowerment” and “intrinsic sustainability”. The core contributions of this study are as follows: it constructs a three-dimensional evaluation framework for a “sustainable intelligent future”; it reveals the dialectical logic of AI as a “double-edged sword”; and it proposes a two-tier governance system comprising a “common framework + differentiated strategies”. The study therefore argues that achieving a sustainable intelligent future requires transcending the instrumental rationality of efficiency-first and establishing a systemic framework that encompasses agile governance, ethical design, and inclusive development.

Keywords: Artificial Intelligence, Sustainable Development, Education, Business, Wellness

引言

人类文明的兴衰更迭，在本质上与技术发展的脉搏同频共振。从工业革命中蒸汽机的轰鸣与电气时代的曙光，到信息技术革命催生的全球互联，每一次技术范式的跃迁都深刻地重塑了社会的生产力结构、组织形态与生活方式。进入二十一世纪，信息化浪潮以排山倒海之势席卷全球，智能手机成为个人与世界交互的神经末梢，智能家居悄然优化着生活的每一处细节，而智能机器人则日益频繁地出现在商场导引、餐厅服务、银行咨询乃至医院分诊等场景中，预示着一种更具沉浸感与便捷性的智能社会图景正逐渐展开。自 2016 年以来，全球围绕人工智能的研究呈爆炸式增长，这项诞生于上世纪中叶的技术，已经成为 21 世纪最受关注的前沿科技之一（梁冠宇, 2021）。

世界各国都已认识到人工智能的战略重要性。以中国为例，2017 年发布的《新一代人工智能发展规划》明确提出，要将人工智能发展提升到国家战略高度。这份文件规划了中国在人工智能领域成为世界强国的路线图，体现了国家对这一技术的重视。2019 年，习近平总书记进一步强调，推动人工智能发展是实现科技跨越的关键。《人工智能发展报告 2024》指出，人工智能正以前所未有的速度改变着生产和生活方式，我们正在迈向通用人工智能的新时代（黄晓琪，2025）。

在这场技术变革中，人工智能不仅仅是提升效率的工具，更是推动社会进步的重要力量。特别是在与人们生活息息相关的教育、商业和健康三个领域，人工智能的融合应用展现出巨大潜力。在教育方面，AI 技术能够根据每个学生的特点提供个性化的学习方案，让“因材施教”真正成为可能。在商业领域，AI 不仅帮助企业提升运营效率，更催生出许多新的商业模式和增长点。在健康领域，AI 在疾病诊断、药物研发和健康管理等方面的应用，为提高医疗水平带来了新的希望。

然而，当前的实践与学术讨论，在很大程度上仍被禁锢于人工智能提升特定领域效率、降低成本的“工具性价值”叙事之中。这种视角的局限性在于，它倾向于将技术视为中性的工具，而忽视了其嵌入社会系统后所引发的系统性、伦理性与长期性影响。人工智能的深度融合，在带来教育个性化、产业智能化、医疗精准化等美好前景的同时，也正在激发一系列严峻且相互关联的挑战：数据隐私与算法黑箱侵蚀着个人自主与社会信任的根基；算法偏见与歧视可能在教育、招聘、信贷、诊疗等关键环节固化甚至加剧社会不平等；数字鸿沟的扩大风险使得技术红利可能仅为少数群体所享有，背离包容性发展的初衷；而支撑人工智能庞大算力需求的巨额能源消耗，则与其本应促进的绿色、低碳可持续发展目标构成了刺眼的悖论。这些挑战超越了单一的技术或管理范畴，直指技术治理、社会公平与全球合作的深层次命题。

正因如此，本研究希望通过系统的分析，达成三个主要目标：第一，深入探讨人工智能在教育、商业和健康三个领域的具体应用方式及其带来的核心价值；第二，全面分析在这些应用过程中可能出现的共同技术风险和社会伦理问题；第三，在此基础上，探讨如何通过多方协作和有效治理，引导人工智能技术朝着更加安全、公平、可持续的方向发展，真正服务于建设一个智能、包容、可持续的未来社会。

本研究将从研究目的、文献综述、研究方法、研究结果、讨论、总结和建议等多个维度，系统地分析人工智能在这些领域的应用模式、面临的风险与挑战，并探讨其未来的发展趋势和实践路径。

研究目的

1. 系统梳理与整合：全面梳理人工智能在教育、商业及健康领域的现有融合应用模式，整合分散的研究成果，绘制 AI 赋能可持续发展的跨领域知识图谱。

2. 深度辨析价值与风险：超越技术效率层面，深入辨析 AI 融合在促进教育公平、驱动绿色商业、实现普惠健康等方面的核心价值，同时系统识别并分析其在数据安全、算法公正、能源消耗及社会包容性等方面引发的风险与悖论。

3. 构建分析框架并提出路径：基于价值与风险的辩证分析，尝试构建“技术赋能-风险规制-可持续治理”的分析框架，并提出促进人工智能健康、公平、可持续地融入关键社会领域的策略与路径，为相关决策与实践提供参考。

文献综述

本章节将从人工智能在教育、商业和健康这三大领域的应用研究与共性挑战研究两个维度进行综述。

1. 分领域融合应用研究

1.1 教育领域：从技术应用到形态重塑

教育领域的人工智能研究呈现出从“技术应用探索”向“教育形态重塑”演进的研究脉络。

早期研究（2018-2020 年）主要聚焦于人工智能教育应用的技术形态与教学变革。曹培杰（2018）提出智慧教育是以“人的智慧成长”为导向，运用 AI 技术推动教育系统性变革，奠定了智慧教育研究的理论基调。梁迎丽和刘陈（2018）系统梳理了智能导师系统（ITS）、自动化测评、教育游戏和机器人四大应用形态，总结出智能化、自动化、个性化、多元化与协同化五大特征。这一时期的研究重在回答“AI 能做什么”的问题，确立了知识图谱（李振等，2019）、智慧课堂（刘邦奇，2020）等核心技术路径。

近期研究（2022-2026 年）则转向“AI 如何重塑教育形态”的深层探讨。胡小勇等（2022）从需求、愿景、路径三方面系统阐述了人工智能赋能教育高质量发展的框架，将研究视野从技术应用拓展至教育公平、智能治理等宏观议题。以 ChatGPT 为代表的生成式人工智能出现后，研究进一步聚焦于新一代 AI 技术的教育变革潜力。卢宇等（2023）基于生成式 AI 的启发性内容生成、对话情境理解、序列任务执行与程序语言解析四项核心能力，分析了其在教学、学习、评价与辅导四个维度的应用可能。刘邦奇等（2024）进一步指出，生成式 AI 将推动教育主体关系、环境、资源、教学方式及评价体系的系统性变革。杨正等（2026）系统梳理了大语言模型、多模态分析等核心技术，强调 AI 正推动教育向个性化、精准化与终身化转型。

1.2 商业领域：从宏观效应到产业融合

商业领域的研究体量最大，研究视角最为多元，呈现出从“宏观经济效应”向“产业深度融合”演进，并进一步向“具体场景落地”拓展的研究脉络。

第一条主线：宏观层面的经济影响与理论阐释。 早期研究主要关注 AI 对宏观经济系统的整体影响。曹静和周亚林（2018）系统综述了 AI 对生产率、经济增长、劳动力就业及收入不平等的作用机制，揭示了 AI 在替代劳动与创造新岗位、加剧收入差距等方面的复杂影响。何玉长和方坤（2018）从理论层面阐释了 AI 作为先进生产力与实体经济深度融合的必然性，指出 AI 通过渗透劳动力、工具、对象等生产力要素，推动制造业转型升级。任保平

和宋文月（2019）进一步提出，新一代 AI 是实体经济高质量发展的新引擎，在改造传统产业、催生新业态、优化要素配置等方面发挥关键作用。郑世林等（2023）以 ChatGPT 为案例，从经济社会视角分析了 AI 带来的劳动力市场极化、教育方式重塑等深层影响。这一主线为理解 AI 的商业价值提供了宏观理论基础。

第二条主线：中观层面的产业融合与智能化转型。 随着研究深入，学者们开始关注 AI 与具体产业的深度融合机制。高煜（2019）聚焦制造业，分析了 AI 在产品、装备、生产、管理、商业和生态六个方面的智能化变革作用。陈亮（2021）探讨了智能制造背景下智慧物流供应链的建设，分析了 AI 在提升物流效率、降低成本、实现精准化服务方面的应用。黄旭和洪美玲（2024）构建了“新质生产力—产业升级—数字生态”三位一体分析框架，系统阐释了生成式 AI 在微观、中观、宏观三个层面对数字经济高质量发展的驱动作用。陈柳钦（2025）则聚焦低空经济这一新兴领域，揭示了 AI 在物流、农业、城市交通等场景中通过智能调度、路径优化与自动化作业催生新商业模式的作用机制。这一主线将研究从宏观效应下沉到产业层面，揭示了 AI 赋能实体经济的具体路径。

第三条主线：微观层面的场景应用与商业模式创新。 近年来，研究进一步聚焦于 AI 在具体商业场景中的落地应用。喻国明（2017）以传媒业为场景，系统分析了 AI 在信息采集、新闻制作、用户体验和内容推送四大环节的变革性应用。屈娟娟（2020）以连锁酒店为例，探讨了 K-means 聚类、决策树、RFM 模型等技术在数字营销中的具体应用，展示了智能营销在客户细分与个性化推荐方面的实践价值。丁志勇和钱楚璇（2026）聚焦商业银行数字化转型，系统阐述了 AI 通过运营自动化、渠道多元化和决策智能化重塑银行价值链的战略路径。这一主线将研究从产业层面下沉到企业运营层面，关注 AI 如何重塑具体业务流程与商业模式。

三条主线并非时间上的先后替代，而是并行演进、相互支撑：宏观理论为产业融合提供分析框架，产业研究为场景落地指明方向，场景实践又反过来丰富和完善宏观理论。

1.3 健康领域：从技术验证到系统治理

健康领域的研究呈现出从“技术应用验证”向“系统治理构建”演进的研究脉络。

前期研究（2017-2021 年）主要聚焦于 AI 技术在医疗场景中的应用潜力与初步问题识别。高奇琦和吕俊延（2017）系统梳理了 AI 在智能诊断、智能治疗、日常健康管理及人性化医疗等方面的应用潜力，同时指出其面临隐私保护、伦理监管、社会观念及就业冲击等挑战。王海星等（2018）进一步将应用场景细化为医疗影像、辅助诊断、健康管理、疾病预测及药物研发等，并系统分析了数据质量、算法歧视、责任界定、法规滞后等现实问题。邢珍珍（2021）聚焦社区智慧养老，构建了融合环境监测、智能分析与终端交互的系统模型，将研究视角从医院场景拓展至社区与家庭场景。这一时期的研究重在验证 AI 在健康领域的技术可行性，并初步识别了应用过程中的风险点。

近期研究（2023-2025 年）则转向“如何构建可持续的健康 AI 治理体系”。周慎等（2020）以新冠疫情防控为例，从危机管理全周期视角系统阐释了 AI 在疫情风险预警、传染源追踪、辅助诊断、事后复盘等环节的赋能效用，将研究视野从技术应用拓展至公共卫生应急管理。朱文珍（2025）聚焦以 Transformer 架构为代表的 AI 大模型，系统梳理了大模型在精准诊疗、药物研发及医学教育三大核心场景的突破性进展。周小芹等（2025）则更全面地概述了 AI 在医学领域的多维度应用，并系统梳理了算法透明度、数据隐私、监管缺失等核心挑战，明确提出未来方向在于提升算法可解释性、构建数据共享机制、完善监管体系及加强跨学科人才培养。

健康领域研究的演进逻辑同样清晰：从早期的“技术可行性验证”，逐步转向“应用风险识别”，再进一步转向“治理体系构建”。这一演进反映了健康领域对 AI 技术审慎、稳健的应用态度，也体现了医疗场景对安全性与伦理性的特殊要求。

2. 共性挑战

随着 AI 在教育、商业和健康领域的深入应用，学者们识别出一系列跨越领域的共性挑战。这些挑战可从三个维度加以整合：

第一，技术维度的风险。 各领域研究均指出，算法黑箱与可解释性不足是制约 AI 深度应用的核心障碍。在教育领域，AI 模型的可解释性不足制约了其在教学评价等关键场景中的可信应用（卢宇等，2023）；在商业领域，AIGC 技术的算法黑箱与行业标准缺失阻碍其健康发展（许雪晨等，2023）；在健康领域，算法透明度不足与诊断依据的可追溯性缺失是核心挑战（周小芹等，2025）。数据安全与隐私泄露风险同样是三个领域共同关注的焦点（王海星等，2018；许雪晨等，2023；梁冠宇，2021）。

第二，社会维度的冲击。劳动力替代与就业结构重塑是跨越三个领域的共性议题。商业领域关注 AI 对中等技能脑力劳动的替代（郑世林等，2023），健康领域则关注 AI 对医护人员的就业冲击（高奇琦和吕俊延，2017）。数字鸿沟与智能不平等问题同样在各领域均有体现（魏钰明等，2025）。

第三，伦理与治理维度的挑战。王军（2018）从哲学层面探讨了 AI 的道德主体地位问题；张成岗（2018）从社会学视角分析了 AI 的自主性风险与责任主体缺失困境；魏钰明等（2025）则指出技术迭代过快引发的治理滞后与社会秩序失灵风险。这些研究共同指向一个核心问题：在 AI 深度嵌入社会系统的背景下，如何构建适应性治理框架以应对技术发展的不确定性。

总体而言，现有研究已充分揭示了 AI 技术在教育、商业、健康领域赋能变革的具体路径与潜在价值，并对数据、算法、伦理、社会影响等跨领域共性风险形成了较为全面的认知。然而，多数研究局限于单一领域内的技术应用或单一维度的风险讨论，缺乏将三大领域置于“可持续智能未来”统一愿景下的系统性、比较性研究，尤其缺乏对“技术赋能”与“可持续发展目标”之间复杂关联与内在张力的深入剖析。本研究旨在弥补这一不足。

研究方法

本研究采用系统性文献综述与内容分析相结合的研究方法，以全面、深入地探讨人工智能在教育、商业与健康等领域的融合发展。

1. 文献检索与筛选

本研究以中国知网（CNKI）为主要文献来源。检索时间范围设定为 2016 年 1 月至 2026 年 2 月。在中国知网中，采用高级检索方式，检索字段选择“篇名”，采用分组检索策略。

针对教育领域，检索条件为：篇名含“人工智能”，且篇名含“教育”。针对商业领域，由于“人工智能+商业”初始检索数量较少（实际检索为 260 篇），为保证样本充分性，将检索范围扩展至“商业”与“经济”两个相关领域。最终检索条件为：篇名含“人工智能”，且篇名含“商业”或“经济”。针对健康领域，由于“人工智能+健康”初始检索数量较少（实际检索为 649 篇），为保证样本充分性，将检索范围扩展至“健康”与“医疗”两个相关领域。最终检索条件为：篇名含“人工智能”，且篇名含“健康”或“医疗”。三组检索分别获得：教育领域 8415 篇，商业与经济领域 1452 篇，健康领域 1709 篇，共 11576 篇。

由于初始文献数量庞大，本研究采用多阶段分层筛选策略。

第一阶段（来源与质量筛选）：将文献来源限定为北大核心期刊、CSSCI 来源期刊。该阶段筛选后，保留 1937 篇。

第二阶段（标题与摘要筛选）：对 1937 篇文献进行标题与摘要阅读，筛选标准包括：（1）研究主题聚焦人工智能在教育、商业或健康领域的具体应用；（2）文献类型为学术论文或研究报告；（3）研究内容涉及应用模式、价值分析、风险识别或治理策略。排除会议综述、书评、新闻报道等。此外一些非核心期刊的一些文献具有参考价值，也纳入其中。该阶段保留 186 篇。

第三阶段（全文筛选与质量评估）：对 186 篇文献进行全文精读，进一步评估相关性、学术质量与代表性。最终选取发表于核心期刊、具有较高被引频次或来源于权威机构报告的代表性文献作为分析基础，共计纳入文献 31 篇，其中教育领域 10 篇（32%），商业与经济领域 12 篇（39%），健康领域 6 篇（19%），综合治理领域 3 篇（10%）。

2. 内容分析

对纳入的 31 篇文献进行定向内容分析，系统识别并归纳人工智能在教育、商业与健康领域中的具体应用模式、技术原理、创新点、面临的挑战（涵盖技术、伦理、社会等维度）以及未来发展趋势。

具体分析过程如下：首先，对每篇文献进行通读，提取与本研究相关的核心信息，包括人工智能在各领域的具体应用形态，例如教育领域的智能导师系统、自动化测评、知识图谱与智慧课堂等，商业领域的智能营销、智能制造与智慧物流等，健康领域的辅助诊断、健康管理与药物研发等；同时提取文献论述的人工智能赋能价值，例如效率提升、个性化服务、资源优化配置等；识别文献指出的风险挑战，例如数据隐私、算法偏见、数字鸿沟等；并梳理文献提出的治理建议。其次，将提取的信息按照领域分类整理，形成教育、商业、健康三个领域的分析材料。最后，对三个领域的分析结果进行横向比较，归纳跨越不同领域的共性议题与各领域的差异特征。

为确保分析的客观性，文献阅读与信息提取由两名研究者独立进行，完成后对提取结果进行交叉核对，对于存在分歧的部分通过讨论达成一致。

3. 综合分析

在内容分析的基础上，本研究采用归纳与演绎相结合的方式对研究发现进行整合。一方面，从分领域的分析材料中归纳出人工智能融合发展的共性模式与挑战；另一方面，以“技术赋能-风险规制-可持续治理”为分析框架，对人工智能在教育、商业和健康领域的融合发展进行系统性审视，探讨构建可持续智能未来的路径与策略。

研究结果

本研究通过对 31 篇核心文献的系统梳理与分析，发现人工智能在教育、商业与健康三大领域的融合应用已形成清晰的实践路径，并展现出多维度的变革性影响，但其发展进程中也浮现出不容忽视的共性风险。各领域因核心任务、技术逻辑与价值取向的差异，呈现出差异化的赋能模式与演进特征。

1. 教育领域：从“规模化教学”到“个性化学习”的范式变革

人工智能正驱动教育系统从统一供给模式向精准服务模式转型。与商业领域的“效率优先”逻辑不同，教育领域的技术融合始终围绕“人的成长”这一核心目标，强调技术对个性化学习与教育公平的支撑作用。

1.1 教学流程的智能化重构

在微观教学层面，AI 通过自适应学习平台与智能导师系统实现了学习路径的动态规划与实时反馈。基于知识图谱的学习系统能够精准诊断学生的知识薄弱点，并提供个性化的学习资源与补救措施（李振等，2019）。研究显示，智能导师系统能够帮助学习者更高效地掌握技能，如美国国防部资助的数字导师系统将海军新兵训练成为技术技能专家所需的时间从几年缩短至几个月（梁迎丽和刘陈，2018）。刘邦奇（2020）以智慧课堂为例构建了“新三段十步”智能教学流程，该流程具有结构合理、操作简便、技术可行的特点。

1.2 教育形态的人机协同转型

生成式人工智能的出现正在重塑教学互动与内容创造方式。卢宇等（2023）以 ChatGPT 为例的研究表明，生成式 AI 能够作为“智能学伴”进行启发式问答，协助教师生成教案、设计习题、批改主观题。刘邦奇等（2024）以星火大模型为例，指出生成式人工智能能够减轻教师日常工作负担，并基于学生特征提供定制化的学习支持。这一转变预示着教育形态正从“教师主导”向“人机协同”演进。

1.3 教育治理的数据化支撑

在系统生态层面，AI 技术通过大数据分析与智能管理工具赋能教育治理现代化。胡小勇等（2022）指出，大数据与人工智能技术能够支持建立教育过程实时监测与智能预警系统，以数据驱动科学决策，提高教育决策的科学性与预见性。这一应用在促进教育公平、实现资源精准投放方面展现出独特价值。

2. 商业领域：从“效率优化”到“业态重塑”的价值跃迁

人工智能在商业领域的融合已超越自动化工具范畴，成为驱动产业智能化升级与价值创造的核心引擎。与教育领域不同，商业领域的技术融合遵循“效率逻辑”，强调对生产力要素的重构与价值链条的再造。

2.1 生产力要素的智能化重构

在微观生产要素层面，AI 通过渗透至劳动力、劳动工具与劳动对象，催生新职业并创造智能产品与服务新业态。何玉长和方坤（2018）的理论研究表明，AI 作为先进生产力要素，能够同时改造劳动力结构、升级劳动工具、拓展劳动对象，形成“三位一体”的变革效应。郑世林等（2023）的研究进一步指出，ChatGPT 等生成式 AI 正在催生提示工程师、AI 训练师等新职业，预计未来五年将创造约 500 万个相关就业岗位。

2.2 产业链条的智能化贯通

在中观产业层面，AI 已实现从研发、生产到流通、销售的全链条赋能。高煜（2019）对制造业的研究表明，AI 在产品、装备、生产、管理、商业应用、产业生态六个维度同时发挥作用，能够显著提升制造业的生产效率。陈亮（2021）对智慧物流供应链的研究显示，AI 技术在路径优化、仓储管理、需求预测等环节的应用，能够有效降低物流成本、提升配送时效。

2.3 新业态的催生与价值创造

在宏观业态层面，AI 正催生新的经济形态与商业模式。黄旭和洪美玲（2024）构建的“新质生产力—产业转型升级—数字生态”三位一体分析框架，系统阐释了生成式 AI 在

微观、中观、宏观三个层面推动数字经济高质量发展的作用机理。陈柳钦（2025）对低空经济的研究进一步揭示，AI 与无人机、物联网技术的融合，正在物流配送、城市交通、农业植保等领域催生全新的商业模式。

3. 健康领域：从“经验医疗”到“精准健康”的模式演进

健康领域的 AI 融合遵循“安全逻辑”，以提升诊疗精准性与患者安全为核心目标，强调技术对临床决策的辅助作用。与商业领域追求效率最大化不同，健康领域对技术可靠性与可解释性有着更为严格的要求。

3.1 临床诊疗的智能化辅助

在临床层面，AI 正从辅助工具向临床决策支持系统演进。高奇琦和吕俊延（2017）系统梳理了 AI 在乳腺癌、肺癌、皮肤癌等癌症诊断领域的应用，并提供了具体数据说明 AI 诊断准确率已超过人类医生。朱文珍（2025）系统梳理了 AI 大模型在医学影像分析、精准诊疗等领域的应用进展，指出大模型能够有效提升早期病灶的检出率。王海星等（2018）指出，人工智能在医学影像识别中的应用可辅助医生阅片，提高工作效率。

3.2 健康服务的智能化延伸

在服务层面，AI 正推动健康管理从医院向社区和家庭延伸。邢珍珍（2021）构建的社区智慧养老系统模型，融合深度学习与物联网技术，能够实时监测老年人健康状况，并实现跌倒检测、用药提醒等功能。

3.3 药物研发的智能化加速

在研发层面，AI 正大幅缩短新药发现周期并降低研发成本。朱文珍（2025）指出，传统新药研发通常耗时 10 年左右、投入约 10 亿美元，而 AI 大模型的应用有望显著提升药物研发效率。这一转变标志着健康领域正在迈向更精准、更普惠、更高效的智能化发展阶段。

4. 共性风险的内在关联与系统性分析

尽管人工智能在三大领域的融合展现出广阔前景，但同时也面临一系列严峻的共性风险。这些风险并非孤立存在，而是呈现出从技术根源到社会外溢再到治理滞后的层次性关联结构。

4.1 技术根源：数据与算法风险

数据与算法层面的风险是共性风险的根源，也是最核心的挑战。在商业领域，许雪晨等（2023）的研究指出，AIGC 技术存在算法黑箱、技术不稳健与行业标准缺失等固有问题。在健康领域，周小芹等（2025）强调算法透明度不足与诊断依据的可追溯性缺失，是制约 AI 医疗深度应用的核心障碍。这些技术缺陷构成了其他风险的触发条件。

4.2 社会外溢：就业冲击与数字鸿沟

技术风险向社会的扩散，形成了第二层次的社会结构风险。郑世林等（2023）的研究表明，AI 对中等技能脑力劳动的替代效应已开始显现，从事软件技术、新闻媒体、法律、金融分析、会计等职业的劳动者面临较大挑战。魏钰明等（2025）进一步指出，智能应用下沉可能扩大数字鸿沟与社会不平等，使技术资源匮乏的群体进一步被边缘化。值得注意的是，这一风险在三个领域表现各异：商业领域直接体现为就业替代，教育领域表现为教师角色转型，健康领域则体现为医护人员的职能重构。但无论形式如何，其内在机制均是技术对传统劳动分工的冲击。

4.3 治理滞后：制度建设的系统性缺失

治理体系层面的滞后是第三层次风险，也是应对前两类风险的关键所在。张成岗（2018）从社会学视角指出，技术迭代速度远超过法律与伦理规范的建设步伐，形成了显著

的“治理赤字”。魏钰明等（2025）进一步揭示，DeepSeek 等大模型的快速迭代已引发治理滞后与社会秩序失灵风险。具体而言，责任认定困境（如 AI 医疗事故责任归属）、安全标准缺失、行业监管空白等问题在各领域普遍存在（王海星等，2018；许雪晨等，2023）。王军（2018）则从哲学层面指出，AI 是否具备道德主体地位这一根本问题尚未厘清，导致责任归属的伦理基础悬而未决。

4.4 风险关联的系统性分析

三类风险呈现出清晰的层级递进关系，这意味着治理层面的风险既是前两类风险未能得到有效管控的结果，也是解决前两类风险的关键切入点。若无法建立健全的治理框架，技术风险将持续外溢为社会风险；反之，若能通过制度创新实现适应性治理，则可有效规制技术风险，缓冲社会冲击。

讨论

本研究的发现揭示了人工智能与可持续发展之间复杂而深刻的互动关系。这种关系并非简单的“工具——目标”线性促进，而是一个充满张力、机遇与风险并存的动态过程。以下将从三个层面展开核心讨论：

1. 对“可持续智能未来”核心内涵的再界定

“可持续智能未来”的核心内涵，已从单纯的技术效能提升，拓展为一个多维、动态且高度整合的价值范式。它要求我们在推进人工智能与经济社会深度融合的过程中，必须超越“效率至上”的单一逻辑，转而将“可持续性”作为衡量一切智能技术应用发展的终极标尺与方向指引。这意味着，一个真正具备可持续性的智能未来，应致力于在社会、经济和环境三大支柱上取得系统性平衡。基于本研究对 31 篇核心文献的系统分析，可初步构建一个包含三个维度、九项核心指标的评估框架，为“可持续智能未来”提供可操作化的评判标准（见表 1）。

表 1: “可持续智能未来”三维度评估框架

维度	核心目标	评判指标
社会维度	公平与包容	1. 数字鸿沟缩小程度：不同地区、群体间的 AI 接入率差距 2. 算法偏见纠正率：算法审计发现的歧视性结果修正比例 3. 公众信任度：用户对 AI 系统可解释性的满意度
经济维度	包容性增长与韧性	4. 就业转型支持率：被 AI 替代岗位的再就业培训覆盖率 5. 财富共享度：AI 效率红利向中小企业与劳动者的溢出效应 6. 产业创新活力：AI 驱动的产业升级对 GDP 的贡献率
环境维度	绿色可持续	7. 能效水平：大模型训练与运行的能耗强度（每百万 token 能耗） 8. 碳足迹控制：AI 算力基础设施的可再生能源使用比例 9. 绿色赋能效应：AI 在能源优化、气候应对中的应用成效

这一框架的意义在于：它将“可持续智能未来”从抽象愿景转化为可测量、可评估、可问责的治理对象。对于政策制定者而言，可据此监测技术发展的社会影响；对于技术开发者而言，可据此在产品设计阶段嵌入价值考量；对于公众而言，则获得了监督 AI 向善发展的透明依据。只有将可持续性量化为具体的评判标尺，我们才能真正超越口号，在实践中引导智能革命走向人本、包容与绿色的未来。

2. 机遇与风险的辩证统一：AI 作为“矛盾复合体”

研究结果清晰地表明，AI 在同一领域中往往扮演着“赋能者”与“挑战者”双重角色，展现出其与生俱来的“双刃剑”效应。这种效应根植于技术逻辑与社会结构之间的复杂互动，而非单纯源于技术本身的缺陷。以下结合本研究的具体发现，剖析三大领域中的典型案例。

在教育领域，AI 驱动的个性化学习平台承载着促进教育公平的美好愿景。李振等（2019）的研究表明，基于知识图谱的学习系统能够精准诊断学生的知识薄弱点，并提供定制化的练习与讲解，为规模化个性化教育提供了技术可能。刘邦奇（2020）以智慧课堂为例构建的“新三段十步”智能教学流程，将因材施教的理念推向了可操作化的新阶段。然而，技术赋能的另一面同样不容忽视。卢宇等（2023）的研究指出，生成式 AI 系统在中文语境中的理解和表达能力总体上弱于英文，可能导致非英语母语的学习者理解困难甚至理解错误。更为根本的挑战在于，优质教育资源的智能化转型，往往率先在发达地区、重点学校落地，而偏远地区、薄弱学校因硬件配置、网络接入、师生数字素养等多重限制，可能在这场智能化浪潮中进一步掉队。初衷在于普惠的技术，若缺乏配套的资源倾斜与能力建设，反而可能将原有的“教育鸿沟”升级为更深层的“智能鸿沟”。

在商业领域，AI 技术的核心价值在于提升效率与优化决策。何玉长和方坤（2018）的理论研究表明，AI 作为先进生产力要素，能够同时改造劳动力结构、升级劳动工具、拓展劳动对象，形成“三位一体”的变革效应。高煜（2019）对制造业的研究进一步揭示，AI 在产品、装备、生产、管理、商业应用、产业生态六个维度的同时作用，可显著提升生产效率。然而，效率提升的背面是就业结构的深刻调整。郑世林等（2023）的研究明确警示，ChatGPT 等生成式 AI 对中等技能脑力劳动的替代效应已开始显现，从事软件技术、新闻媒体、法律、金融分析等职业的劳动者面临较大挑战。陈亮（2021）对智慧物流的研究也指出，AI 技术在提升效率的同时，也对传统物流岗位产生替代效应。效率的追求若不与公平的考量相结合，可能带来结构性失业与社会矛盾。正如许雪晨等（2023）所指出的，AIGC 技术带来的算法黑箱与行业标准缺失，也可能阻碍其健康发展。

在健康领域，AI 辅助诊断系统为提高诊疗效率与精准度带来了巨大希望。高奇琦和吕俊延（2017）的研究指出，AI 在癌症识别等领域的诊断准确率已接近甚至超过资深医师。朱文珍（2025）系统梳理的 AI 大模型应用也显示，大模型在医学影像分析、药物研发等领域展现出突破性潜力。然而，技术本身的中立性并不保证应用结果的公平性。王海星等（2018）的研究明确指出，医疗 AI 面临数据质量、算法歧视、责任界定等核心挑战。若训练数据未能充分代表多样化的患者群体（如特定种族、性别或年龄组），模型便可能对这些群体产生系统性误诊或疗效预测偏差。周小芹等（2025）也强调，算法透明度不足与诊断依据的可追溯性缺失，是制约 AI 医疗深度应用的核心障碍。旨在拯救生命的技术，若忽视社会语境，反而可能造成新的健康不平等。

上述三大领域的案例深刻揭示了，技术并非在真空中运行，其影响由其设计逻辑、部署环境、治理框架与社会价值导向共同塑造。因此，我们必须摒弃技术决定论式的乐观或悲观，既不能指望技术自动解决一切社会问题，也不应因潜在风险而全然抗拒进步。相反，需要转向一种辩证且系统性的治理思维：在推动 AI 创新与应用的同时，主动预见、评估并管理其可能引发的社会、伦理与经济风险；在技术设计中嵌入公平、透明、可问责的价值考量。

3. 跨领域共性挑战与差异化治理路径

人工智能在医疗、教育、商业等领域的深入融合，持续暴露出数据隐私、算法偏见、责任界定等共性挑战。然而，这些共性挑战在不同领域呈现出不同的风险特征与治理逻辑。因此，未来的治理核心，在于共性框架与差异化策略的结合：一方面，需要构建跨领域协同的基础治理体系；另一方面，必须针对各领域的特殊性，设计差异化的治理路径。

3.1 共性治理框架

三大领域面临的共性挑战表明，其性质已超越单一行业的应用问题，上升为嵌入技术社会底层的结构性风险。传统的条块分割、被动回应式监管已难以应对 AI 技术的泛在渗透与快速迭代。为此，需构建一个跨领域协同、敏捷适应、贯穿技术全生命周期的综合性治理框架，包含三个核心机制：一是伦理前置机制。将公平、透明、可问责等价值原则内嵌于技术研发早期阶段，而非事后补救。二是多层次治理网络。融合法律规制、行业技术标准、企业自律机制与公众参与监督，形成刚柔并济的约束体系。三是前瞻性治理工具。利用技术影响评估、社会风险模拟、沙盒测试等工具，主动识别与防范潜在风险。

3.2 差异化治理策略

在共性框架之上，各领域因其核心任务、技术逻辑与价值取向的差异，需要设计差异化的治理策略。

教育领域注重能力建设与普惠保障。教育领域的核心价值是“育人”与“公平”。治理应聚焦于三个方面：一是将师生智能教育素养纳入培训体系；二是建立教育资源智能化的区域补偿机制，防止“技术鸿沟”转化为“教育鸿沟”；三是针对生成式 AI 引发的学术伦理失范问题，建立明确的使用规范与学术诚信审查机制。

商业领域注重竞争秩序与就业转型。商业领域的核心逻辑是“效率”与“创新”。治理应聚焦于三个方面：一是针对“算法黑箱”问题建立算法审计与信息披露制度；二是针对中等技能脑力劳动替代效应，建立终身学习体系与职业转型支持机制；三是明确 AI 系统开发者、部署者、使用者的权责边界。

健康领域注重安全监管与伦理审查。健康领域的核心逻辑是“安全”与“精准”。治理应聚焦于三个方面：一是建立医疗 AI 数据的全生命周期安全管理机制；二是建立医疗 AI 产品的临床准入标准与效果评估体系，确保算法可追溯性与问责机制；三是在医疗 AI 研发中引入独立的伦理审查委员会，预评估算法偏见风险。

共性挑战与差异化治理的统一，揭示了人工智能治理的核心命题：既需要统一的价值观引领与跨领域协同的基础设施，又必须尊重各领域的特殊逻辑，将治理原则转化为场景适配的具体制度设计。唯有通过“共性框架+差异化策略”的双层治理体系，才能有效驾驭 AI 的双重属性，确保其沿着促进社会公平、经济包容与环境可持续的轨道前进。这不仅是技术管理的升级，更是关乎未来社会形态的深刻制度创新。

总结

本研究系统探讨了人工智能融入教育、商业与健康领域以构建可持续智能未来的图景、路径与挑战。研究表明，AI 通过驱动教育个性化、商业智能化和医疗精准化，展现出赋能社会各领域迈向卓越的强大潜力。然而，这一融合进程绝非坦途，其伴生的数据与算法风险、社会结构性风险以及治理滞后风险，构成了对可持续发展的严峻威胁。

研究的核心启示在于，“智能”本身并非天然导向“可持续”。技术赋能的璀璨光芒之下，阴影同样深重。构建真正的可持续智能未来，其本质是一场深刻的社会技术实验，它要求我们以极大的智慧与责任感，在拥抱技术创新与防范社会风险之间审慎前行。这不仅是技术问题，更是关乎价值选择、利益分配与全球协作的治理命题。

建议

基于本研究的分析，为促进人工智能在教育、商业和健康领域的可持续和负责任发展，本研究提出以下建议。

1. 构建敏捷协同的多层次治理体系

为推动人工智能的可持续发展，需建立国家、行业与企业三层协同的治理框架。国家层面应加快出台综合性法律，明确算法透明度、数据权利与责任归属等原则性要求；行业层面需制定细化的技术标准与应用伦理指南，特别是在教育评价、医疗诊断等高风险场景；企业则应设立内部伦理审查机制，将安全与公平指标纳入产品研发全流程。同时，建立跨部门协同监管机制，以应对技术快速迭代带来的监管挑战。

2. 实施贯穿技术生命周期的风险防控

将伦理与可持续发展要求前置融入技术研发环节，推行“伦理设计”与“影响评估”双轨机制。一方面，鼓励研发可解释 AI、联邦学习等提升透明度与隐私保护能力的技术；另一方面，在高风险 AI 系统部署前，强制开展涵盖公平性、安全性及社会影响的系统性评估。特别要关注能源效率，通过推广绿色算法与节能硬件，降低人工智能发展的环境代价。

3. 夯实包容发展的社会基础

采取积极措施防止技术加剧社会分化。应实施包容性数字战略，通过公共投资弥合基础设施与数字素养的鸿沟，确保人工智能服务的普惠可及。建立适应劳动力市场变化的终身学习体系，重点为受自动化冲击的劳动者提供技能转型支持。此外，应加强公众参与，通过科普教育提升全社会对人工智能的理性认知与监督能力，形成技术发展与人文关怀的良性互动。

参考文献

丁志勇, & 钱楚璇. (2026). 人工智能赋能商业银行数字化转型：战略路径、技术融合与生态重构. *新金融,* (1), 53-62.

王军. (2018). 人工智能的伦理问题：挑战与应对. *伦理学研究,* (4), 79-83. https://doi.org/10.15995/j.cnki.llxyj.2018.04.012

王海星, 田雪晴, 游茂, 陆雪秋, 顾泽龙, & 程龙. (2018). 人工智能在医疗领域应用现状、问题及建议. *卫生软科学, 32*(5), 3-9.

卢宇, 余京蕾, 陈鹏鹤, & 李沐云. (2023). 生成式人工智能的教育应用与展望——以 ChatGPT 系统为例. *中国远程教育, 43*(4), 24-51. https://doi.org/10.13541/j.cnki.chinade.20230301.001

邢珍珍. (2021). 人工智能赋能下社区智慧养老服务模式及关键技术研究. *护理研究, 35*(9), 1573-1579.

朱文珍. (2025). 人工智能大模型在医疗健康领域的应用现状与展望. *中国医学影像技术, 41*(8), 1195-1199. https://doi.org/10.13929/j.issn.1003-3289.2025.08.001

任保平, & 宋文月. (2019). 新一代人工智能和实体经济深度融合促进高质量发展的效应与路径. *西北大学学报（哲学社会科学版）, 49*(5), 6-13. https://doi.org/10.16152/j.cnki.xdxbsk.2019-05-001

刘邦奇, 聂小林, 王士进, 袁婷婷, 朱洪军, 赵子琪, & 朱广袤. (2024). 生成式人工智能与未来教育形态重塑：技术框架、能力特征及应用趋势. *电化教育研究, 45*(1), 13-20. https://doi.org/10.13811/j.cnki.eer.2024.01.002

刘邦奇. (2020). 智能技术支持的“因材施教”教学模式构建与应用——以智慧课堂为例. *中国电化教育,* (9), 30-39.

许雪晨, 田侃, & 李文军. (2023). 新一代人工智能技术（AIGC）：发展演进、产业机遇及前景展望. *产业经济评论,* (4), 5-22. https://doi.org/10.19313/j.cnki.cn10-1223/f.20230705.001

李振, 周东岱, & 王勇. (2019). “人工智能+” 视域下的教育知识图谱：内涵、技术框架与应用研究. *远程教育杂志, 37*(4), 42-53. https://doi.org/10.15881/j.cnki.cn33-1304/g4.2019.04.006

杨正, 陶德超, & 孙超. (2026). 人工智能在教育领域的发展与应用. *科技风,* (1), 140-142. https://doi.org/10.19392/j.cnki.1671-7341.202601045

何玉长, & 方坤. (2018). 人工智能与实体经济融合的理论阐释. *学术月刊, 50*(5), 56-67. https://doi.org/10.19862/j.cnki.xsyk.2018.05.006

张成岗. (2018). 人工智能时代：技术发展、风险挑战与秩序重构. *南京社会科学,* (5), 42-52. https://doi.org/10.15937/j.cnki.issn1001-8263.2018.05.006

陈柳钦. (2025). 低空经济与人工智能深度融合驱动新质生产力跃升. *经济与管理研究,* (1), 88-99.

陈亮. (2021). 智能制造背景下智慧物流供应链建设研究. *商业经济研究,* (5), 104-107.

周小芹, 刘慧珍, 王婷, 刘雪婷, 刘芳, & 康德英. (2025). 人工智能赋能医学领域的挑战与发展方向. *中国胸心血管外科临床杂志, 32*(2), 244-251.

周慎, 朱旭峰, & 薛澜. (2020). 人工智能在突发公共卫生事件管理中的赋能效用研究——以全球新冠肺炎疫情防控为例. *中国行政管理,* (10), 35-43. https://doi.org/10.19735/j.issn.1006-0863.2020.10.05

郑世林, 姚守宇, & 王春峰. (2023). ChatGPT 新一代人工智能技术发展的经济和社会影响. *产业经济评论,* (3), 5-21. https://doi.org/10.19313/j.cnki.cn10-1223/f.20230310.002

屈娟娟. (2020). 人工智能及大数据技术在数字营销中的应用. *商业经济研究,* (10), 78-80.

胡小勇, 孙硕, 杨文杰, & 丁格莹. (2022). 人工智能赋能教育高质量发展：需求、愿景与路径. *现代教育技术, 32*(1), 5-15.

高奇琦, & 吕俊延. (2017). 智能医疗：人工智能时代对公共卫生的机遇与挑战. *电子政务,* (11), 11-19. https://doi.org/10.16582/j.cnki.dzzw.2017.11.002

高煜. (2019). 我国经济高质量发展中人工智能与制造业深度融合的智能化模式选择. *西北大学学报（哲学社会科学版）, 49*(5), 28-35. https://doi.org/10.16152/j.cnki.xdxbsk.2019-05-004

黄旭, & 洪美玲. (2024). 生成式人工智能助力数字经济高质量发展的影响机制与提升路径. *南方经济,* (8), 23-44.

黄晓琪. (2025). *人工智能应用背景下媒体暴露对员工失业风险感知的影响研究* [硕士论文]. 南京信息工程大学. https://doi.org/10.27248/d.cnki.gnjqc.2025.001937

曹培杰. (2018). 智慧教育：人工智能时代的教育变革. *教育研究, 39*(8), 121-128.

曹静, & 周亚林. (2018). 人工智能对经济的影响研究进展. *经济学动态,* (1), 103-115.

梁迎丽, & 刘陈. (2018). 人工智能教育应用的现状分析、典型特征与发展趋势. *中国电化教育,* (3), 24-30.

梁冠宇. (2021). *人工智能应用于教育的伦理风险与规避* [硕士论文]. 山西大学. https://doi.org/10.27284/d.cnki.gsxiu.2021.000537

喻国明, 兰美娜, & 李玮. (2017). 智能化：未来传播模式创新的核心逻辑——兼论 “人工智能+媒体” 的基本运作范式. *新闻与写作,* (3), 41-45.

魏钰明, 贾开, 曾润喜, 何哲, 邱林, 于文轩, 唐曼, 黄璜, 曾雄, 张宏, 郑磊, 张会平, 张效羽, 赵静, 傅宏宇, & 蒋余浩. (2025). DeepSeek 突破效应下的人工智能创新发展与治理变革. *电子政务,* (3), 2-39. https://doi.org/10.16582/j.cnki.dzzw.2025.03.001

A COMPREHENSIVE SURVEY ON IMAGE REPRESENTATION AND GENERATION ALGORITHMS FOR MULTI-VIEW AND MULTI-MODAL DATA

Jian Qu[1]*, Somrote Komolavanij[2], Phannachet Na Lamphun[3], Annop Monsakul[4], Kobkrit Viriyayudhakorn[5], Pensiri Manomaisupat[6], and Lin Qian[7]

[1,2,3,4,7]Facutly of Engineering and Technology, Panyapiwat Institute of Management, Nonthaburi, Thailand
[5]iApp Technology Co., Ltd., Pathum Thani, Thailand
[6]O.O.S.Trading & Services Ltd., Bangkok, Thailand
***Corresponding Author, E-mail**: JianQu@pim.ac.th

Abstract

The rapid growth of multimedia data—often involving multiple views (different perspectives of the same object) and multiple modalities (such as images, text, and audio)—is transforming fields like computer vision, medical imaging, and autonomous systems. However, the diverse nature, high complexity, and rich semantics of such data pose major challenges for traditional machine learning methods.

This survey explores the latest advances in learning meaningful representations and generating new content from multi-view and multi-modal data. We begin by introducing key concepts and challenges—such as the semantic gap, data redundancy, and difficulties in aligning and blending different types of data. The survey then systematically examines a wide range of methods, from traditional subspace learning and early deep models to modern transformer-based architectures and diffusion models.

Let's dive into how we represent images. We are talking deep metric learning, self-supervised techniques, and information-theoretic ideas like the Information Bottleneck. Generative AI? We dig into that too—from the early days of GANs and VAEs to the latest Denoising Diffusion Probabilistic Models (DDPMs), plus how these tools handle cross-modal generation. We also break down data fusion strategies—early, mid-level, late fusion—and look at how large-scale vision-language models like CLIP have changed the game, especially when it comes to zero-shot learning.

We still face real challenges with model scalability, making sense of how models work, and getting the most out of our data. There are some exciting research paths ahead, like blending neural and symbolic AI, developing foundation models that handle different types of data, and pushing closer to artificial general intelligence. With this survey, we want to give researchers and practitioners a clear look at where things stand and help them push multi-view and multi-modal data analysis even further.

Keywords: Multi-view Learning, Multi-modal AI, Image Representation, Deep Learning, Generative AI, Diffusion Models, Data Fusion, Vision-Language Models, Information Bottleneck, Cross-modal Retrieval.

Introduction

Big data isn't just about having more information—it's about having all sorts of different information, coming in from everywhere, all at once. Advances in sensors and the internet have turned this trickle into a flood. But here's the thing: a lot of this data isn't straightforward. Take multi-

view data, for example. You're looking at the same thing, but from different angles or sources. Maybe a bunch of cameras capture the same object, or you break down an image using different features like SIFT, HOG, or deep learning. Then there's multi-modal data. That's when you mix different types of information about the same thing. Think of a photo with a caption, or a video with both sound and subtitles. It all adds up to data that's not just bigger—it's richer and more tangled than ever (Baltrusaitis et al., 2019). Bringing together these different sources of information gives us a much fuller, stronger picture than relying on just one view or type of data. That's why you see this approach popping up everywhere—pedestrian re-identification (Zheng et al., 2016) , medical imaging (like blending MRI and CT scans) (Litjens et al., 2017) , visual question answering (Antol et al., 2015) , even self-driving cars (Janai et al., 2020) . Still, working with multi-view or multi-modal data isn't easy. One big headache is the "semantic gap"—it's tough to connect the raw details in the data with the bigger, more abstract ideas we care about. Then there's the issue of high dimensionality and redundancy; all that extra data ramps up the complexity and makes overfitting more likely. Plus, data from different sources often look and behave pretty differently from each other—a problem known as "view heterogeneity" or the "modality gap." That makes it tricky to create a single, unified way of representing everything. And on top of that, there's "cross-modal alignment." Basically, you need to match things up correctly across different types of data—like connecting parts of an image to the right words in a caption. It's a lot to juggle, but that's what makes this field so interesting.

This survey aims to provide a comprehensive and structured overview of the algorithms developed to address these challenges, focusing on two core tasks: image representation learning and image generation. Our contribution is fourfold: Here's what we're doing: laying out a roadmap of methods for learning from multiple views or data types. We cover it all — from old-school models to deep learning and information theory. When it comes to generative models, we really dig in, starting with GANs and VAEs, moving up to newer diffusion models, and showing how these work with different kinds of data. Data fusion? We break that down too, looking at how it works at different stages, and we spotlight how big pre-trained models are changing the game. We don't shy away from the tough stuff either — we talk about what's still hard and where research should head next. Here's how the paper flows: Section 1 is all about learning image representations. Section 2 goes deep into image generation. Section 3 reviews ways to combine different data sources. And in Section 4, we wrap things up and point to what's ahead.

1. Deep Dive into Image Representation Learning

The primary goal of representation learning is to transform raw data into a feature vector that captures its essential characteristics, making it more suitable for tasks like classification, clustering, and retrieval.

1.1 Evolution of Feature Learning Paradigms

The journey of image representation has evolved from manual feature engineering to automated deep feature learning. Back in the early days of computer vision, people spent a lot of time crafting features by hand. SIFT (Lowe, 2004) and HOG (Dalal & Triggs, 2005) are classic examples—they did a solid job picking up on textures and shapes in images. Usually, researchers paired these features with fairly simple models like SVMs (Cortes & Vapnik, 1995) to handle classification. These setups made sense; they were easy to understand and didn't eat up much computing power. But they just couldn't grasp the bigger picture in images, and honestly, you needed a fair bit of specialized knowledge to get good results.

AlexNet's big win at the 2012 ImageNet competition changed everything. Suddenly, deep learning wasn't just a theory — it was the way forward. Convolutional Neural Networks (CNNs) proved they could figure out complex features from raw pixels all on their own. After that, models

like VGGNet (Simonyan & Zisserman, 2014) , GoogLeNet (Szegedy et al., 2015) , ResNet (He et al., 2016) , and DenseNet (Huang et al., 2017) kept pushing the boundaries. They made networks deeper, tackled problems like vanishing gradients, and showed that you really could train massive models on huge datasets. Pretty soon, everyone started using features from these pre-trained CNNs for all sorts of visual tasks. It became the new normal.

More recently, the field has been influenced by two major trends. First, self-supervised learning methods, such as contrastive learning (e.g., SimCLR; Chen et al., 2020) , have emerged to learn powerful some newer approaches don't need huge labeled datasets. Instead, they set up little challenges—pretext tasks—using unlabeled data. The model has to figure out useful patterns on its own. Then there's the Transformer architecture (Vaswani et al., 2017) , which changed the game in language processing and now works for vision too. The Vision Transformer (ViT) (Dosovitskiy et al., 2020) looks at an image as a bunch of patches, kind of like words in a sentence, and uses self-attention to catch details across the whole image. It often beats CNNs when you have a lot of data. Hybrid models like Swin Transformer (Liu et al., 2021) bring back some of the structure from CNNs, making things faster and better.

1.2 Advanced Multi-view Representation Learning Strategies

Back in the day, early multi-view techniques like Canonical Correlation Analysis (CCA) (Hotelling, 1936) tried to find a linear space where two different views lined up as closely as possible. Deep learning took this further with models like Deep Canonical Correlation Analysis (DCCA) (Andrew et al., 2013) , which use neural networks to map each view into a shared space in a nonlinear way. Then came models like Deep Canonical Correlation Autoencoders (DCCAE) (Wang et al., 2015) , which not only maximize the correlation but also try to keep the reconstructed data accurate. This ends up giving more solid representations.

There's also a more theory-driven way to think about this, using information theory. The Information Bottleneck (IB) principle (Tishby et al., 2000) gives a framework for learning features that capture as much as possible about the target you care about—like a class label—while tossing out extra noise from the input. In multi-view setups, you can tweak this idea to keep only what's shared across different views and ignore the stuff that's unique to one side. Deep learning versions of IB, like the Variational Information Bottleneck (VIB) (Alemi et al., 2016) , help models generalize and stay robust by acting as a strong regularizer.

Contrastive learning has found a natural home in multi-view setups. Here's the basic idea: you take different views of the same data—maybe two versions of the same image after different augmentations, or an image paired with its caption—and treat those as positive pairs. Views from different data points? Those are negative pairs. The model learns by pulling the positives closer in embedding space and pushing the negatives apart. This simple trick has worked wonders, especially in cross-modal retrieval. Models like CLIP (Radford et al., 2021) , which got pre-trained on hundreds of millions of image-text pairs, end up with a shared space where related concepts from different types of data cluster together. Table 1 lays out the main contrastive learning-based multimodal models, along with other popular generative and understanding approaches. It breaks things down by modality, main task, strengths, and weaknesses, so it's easier to see how the field has evolved and what problems are still hanging around.

Table 1: Comparison of Multimodal Generation and Understanding Models

Model	Modality Type	Main Task	Advantages	Limitations
Kosmos-1	Text + Image	Multimodal understanding and generation	Strong cross-modal reasoning, high generalization	High computational cost, limited interpretability
GPT-4 (Multimodal)	Text + Image	Multimodal dialogue and content generation	Effective cross-modal alignment, strong transfer ability	Requires large-scale paired data
CLIP	Text + Image	Cross-modal retrieval and matching	Effective cross-modal alignment, strong transfer ability	Requires large-scale paired data
ALIGN	Text + Image	Cross-modal representation learning	Good generalization ability	Sensitive to noisy data
BERT + Vision Encoder	Text + Image	Semantic alignment and multimodal understanding	Fine-grained semantic representation	Complex architecture, high training cost
DALL·E	Text → Image	Text-to-image generation	High creativity, high-quality image synthesis	Limited fine-grained semantic control
DALL·E 2	Text → Image	High-quality image generation	Rich details, strong diversity	Slow inference speed
Stable Diffusion	Text → Image	Image generation and editing	High image quality, controllable outputs	Computationally expensive sampling
GLIDE	Text → Image	Text-to-image generation	Stable generation, flexible style control	Potential semantic drift
DDPM	Image / Text → Image	High-fidelity image generation	Superior image quality compared to GAN/VAE	Low sampling efficiency
Guided Diffusion	Multimodal-conditioned	Conditional image generation	Improved controllability	Increased structural complexity
VQ-Diffusion	Multimodal-conditioned	Cross-modal image generation	Reduced mode collapse, improved stability	Complex training process
GAN	Image / Multimodal	Image generation	Sharp and realistic image synthesis	Training instability, mode collapse risk
VAE	Image / Multimodal	Generative modeling	Stable training, smooth latent space	Blurry image outputs
StyleGAN	Image	High-quality image generation	Fine-grained style control, high realism	High computational cost
MedGAN	Multimodal Medical Imaging	Medical image synthesis	Mitigates data scarcity in healthcare	Clinical reliability and safety concerns

2. Comprehensive Overview of Image Generation Algorithms

Generative models have transitioned from being tools for data augmentation to engines of creative content generation, powering applications in art, design, and synthetic data creation.

2.1 Generative Model Families: A Comparative Analysis

Let's talk about Generative Adversarial Networks (GANs). Goodfellow and his team introduced them (Goodfellow et al., 2014) as a sort of game: a generator tries to make fake data look real, while a discriminator tries to tell the difference between real and fake. The generator takes random noise and turns it into synthetic samples, doing its best to fool the discriminator, which is always on the lookout for fakes. This back-and-forth is what drives both networks to get better—one learns to make better fakes, the other gets sharper at spotting them. GANs made a splash by generating images that looked incredibly sharp and realistic, but training them isn't easy. You run into things like mode collapse, where the generator gets stuck making the same kind of output over and over, and training can be unstable, with losses bouncing all over the place or gradients vanishing.

There's been a lot of work on fixing these problems. For example, the Wasserstein GAN (WGAN) (Arjovsky et al., 2017) switched up the loss function to use Earth-Mover distance and brought in tricks like weight clipping and gradient penalty to make training a lot more stable. Then there's StyleGAN (Karras et al., 2019) , which shook things up with a new generator design. It separates high-level features from random details, thanks to adaptive instance normalization and a mapping network that lets you control image features at different levels of detail.

StyleGAN (Karras et al., 2019) changed the game by letting people control generated images like never before. Its style-based generator splits high-level features from random details using adaptive instance normalization (AdaIN), and its new mapping network turns latent vectors into a style space. That means users get fine-grained control over image features at every resolution.

Variational Autoencoders (VAEs) take a different approach, rooted in probability theory. Here, you have an encoder that maps input data to parameters of a probability distribution in some learned latent space—usually a multivariate Gaussian. Then the decoder takes samples from this distribution and tries to reconstruct the input. Training is all about balancing two things: making sure the reconstructions look like the original data, and keeping the latent space organized using the KL divergence. VAEs generally train more reliably than GANs and give you these nice, smooth latent spaces where you can interpolate between points or do semantic arithmetic. They're also good for things like anomaly detection. But there's a catch: VAE-generated images often look a bit blurry, missing fine details. That's mostly because common loss functions (like mean squared error) focus too much on pixel accuracy and not enough on texture. People have tackled this in a bunch of ways: β-VAE for disentangled features, VQ-VAE with discrete latent codes, and NVAE with hierarchical latent structures. Each one pushes the sample quality further while keeping the core benefits of the VAE framework (Kingma & Welling, 2013) .

Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020) are the new heavyweights in generative modeling. They've pulled off some impressive wins across different fields. The idea is simple but clever: generation happens by slowly removing noise, following two Markov chains. The forward chain just adds more and more Gaussian noise to the data until it's basically random noise. The reverse chain, handled by a neural network (usually a time-aware U-Net), learns to remove this noise step by step, effectively reversing the process. When you want to generate something new, you start with pure noise and run it through the reverse steps to get a fresh data sample.

DDPMs stand out for creating sharp, diverse results that often beat GANs in both visual quality and FID scores. Training is also way more stable, since it turns into a bunch of noise-removal tasks rather than the balancing act GANs require. But there's a catch: generating a single sample takes a long time—hundreds or thousands of steps—since each one needs its own neural network pass. That's a problem for real-time use. To fix this, researchers have come up with faster sampling tricks like DDIM (Song et al., 2020) (which uses fewer steps), knowledge distillation to train smaller networks that mimic the long process, and new schedulers that make the noise removal more efficient. On top of that, people have added new conditioning methods—like classifier guidance, classifier-free guidance, and cross-attention layers—which let diffusion models handle text, images, or almost any other modality for flexible, controllable generation.

2.2 Cross-Modal Generation and Editing

One major use for generative models is conditional generation—basically, creating or tweaking data in one form using signals from another. This bridges the gap between different types of data and opens up new worlds for creativity, data augmentation, and interactive design.

Text-to-Image Generation involves generating a photorealistic or artistic image from a natural language description. This task represents one of the most direct tests of a model's ability to interpret and visualize complex semantics. Early pioneering approaches primarily used Generative Adversarial Networks (GANs) conditioned on text embeddings, where a text encoder would produce a vector that modulated the generator's process. While these models demonstrated proof-of-concept, they often struggled with compositional complexity, coherence in multi-object scenes, and generating high-resolution, diverse outputs.

The field was revolutionized by the advent of large-scale models like DALL·E 2 (Ramesh et al., 2022) and Stable Diffusion (Rombach et al., 2022) . These models leverage the power of diffusion processes and massive, web-scale training datasets. DALL·E 2, for instance, employs a two-stage process: a prior model generates a CLIP image embedding from the text caption, and a decoder model then synthesizes the image from this embedding using diffusion. Stable Diffusion, in particular, is notable for its efficiency and accessibility; it performs the diffusion process not in the high-dimensional pixel space but in a lower-dimensional latent space learned by a Variational Autoencoder (VAE), significantly reducing computational cost and memory requirements while maintaining high quality.

These advanced models demonstrate an impressive ability to understand nuanced semantics, compose novel scenes with multiple entities and specified relationships, and adhere to artistic styles described in the prompt. However, challenges remain in spatial reasoning (e.g., correctly rendering "a cat sitting to the left of a dog"), handling precise counts, and ensuring factual accuracy for complex descriptions.

Early on, people tried using Generative Adversarial Networks (GANs) that were conditioned on text. Basically, they'd take a sentence, turn it into a vector using a text encoder, and then feed that into the generator. These first models managed to show the idea worked, but they fell short when it came to putting together complex scenes, keeping everything coherent when there were lots of objects, or producing high-res, varied images.

Things really took off with the arrival of large-scale models like DALL·E 2 (Ramesh et al., 2022) and Stable Diffusion (Rombach et al., 2022) . These models use diffusion processes and are trained on giant datasets scraped from the web. DALL·E 2, for example, runs in two steps: a "prior" model turns the text into a CLIP image embedding, and then a decoder turns that embedding into an image using diffusion. Stable Diffusion stands out for being fast and accessible.

It doesn't do diffusion in the huge pixel space but works in a smaller latent space learned by a Variational Autoencoder (VAE). This cuts down on memory and compute without sacrificing quality.

Style transfer's been a dream in computer vision for ages: take the look of a Van Gogh painting and put it onto a photo, for instance. Older methods, like the one from Gatys and colleagues (Gatys et al., 2016) , worked by juggling content and style losses in a slow optimization loop. Generative models changed everything. CycleGAN (Zhu et al., 2017) , for example, can learn to swap between two visual styles (like photographs and Monet paintings) even when the images aren't paired up. Once trained, it's fast. More recently, diffusion models have raised the bar again, delivering better quality and more flexibility. Sometimes you don't even need a reference image—just say "make it look like a watercolor," and the model does the rest. With these tools, people can do everything from adding artistic filters to photos to creating realistic training data for machine learning models that need to handle new environments.

3. Data Fusion Methodologies and Large-Scale Models

The strategy for combining information from multiple modalities is critical to the performance of multi-modal systems.

3.1 Fusion Techniques

Early Fusion is a technique where you combine raw data or low-level features from different sources—like text and images—right at the start of feature extraction. The goal is to link up these different types of data early on, so the model can learn richer, more connected representations. People usually do this by sticking features together, blending them with weights, or using more complex tensor-based approaches.

But early fusion isn't easy. The biggest headache is that different data types just aren't alike. Images are made up of continuous pixel values, while text is a sequence of discrete symbols. Trying to combine them directly can be tricky. Another problem is lining everything up—multimodal data needs precise timing or spatial alignment, and even a small mismatch can mess things up. Plus, any noise or errors from one source can bleed into the others during fusion, which can drag down overall performance.

Even with all the obstacles, early fusion still has its place. Take video analysis, for example—it's great at catching those tiny timing overlaps between visuals and audio. In medical imaging, early fusion shines when you need to pull together information from different scans, like CT and MRI, because it helps spot details you'd probably miss if you looked at each one alone.

Mid-level fusion is the main workhorse right now. It lets each type of data—say, images or text—go through its own specialized encoder. Then, once each has been processed, the system brings them together at the feature level. The beauty of this approach is that you get to keep all the unique info from each source, but you also start to find those interesting connections between them.

There's no single way to do mid-level fusion. Sometimes it's as simple as sticking features side by side or running element-wise math. Other times, it gets fancy with dynamic weighting, using attention mechanisms to decide which features matter most, or even tensor fusion for deeper interactions. Attention mechanisms really stand out—they let the model tune feature weights on the fly, depending on what's coming in. In visual question answering, for instance, the system can home in on the parts of an image that actually matter for a given question, lining up visual and language info in a smart way. Mid-level fusion also opens the door to shared representation learning. By designing clever loss functions, you can teach the model to place similar cross-modal content close together in its internal space. This is why mid-level fusion works so well for things like cross-modal retrieval and content generation—it just gets those subtle relationships.

Late fusion takes a different tack. Each type of data does its own thing, extracting features and even making some decisions before everything comes together at the end. This setup brings a couple of big wins. The advantages of this approach are manifold: First, robustness: if one data source is missing or just plain bad, the system can still lean on the others and keep going. That's a lifesaver in real-world situations. Second, you get a lot of flexibility—each modality can use whatever model fits best, and updates are easier. Plus, you can actually see and measure how much each data source is contributing, which helps with transparency.

But late fusion isn't perfect. Because the data streams don't really interact deeply, it struggles with tasks that need to catch subtle cross-modal cues. In something like multimodal sentiment analysis, late fusion might miss the way a person's facial expression and tone of voice work together.

Lately, researchers have been busy finding ways to get around the old problems with traditional fusion methods. They've started mixing different fusion stages, like doing some early fusion at the feature level while keeping the flexibility to blend things together later at the decision stage. Some new methods can actually switch up the fusion approach on the fly, depending on what kind of data comes in and how good it is. There's also this whole wave of meta-learning-based fusion—basically, the system teaches itself how to pick the best fusion weights for whatever job it's facing (Tsai et al., 2019) .

With the development of deep learning technologies, new fusion methods continue to emerge. Memory network-based fusion stores and retrieves cross-modal information through external memory units; Graph neural network-based fusion represents different modalities as graph structures, utilizing graph reasoning to capture complex inter-modal relationships; Adversarial learning-based fusion promotes deeper feature interaction between modalities through generative adversarial networks. All these fresh ideas are making multimodal fusion smarter and much more adaptable, setting the stage for more advanced multimodal AI. As tech like neural architecture search and self-supervised learning keeps improving, we're bound to see even faster, stronger, and more efficient fusion systems pop up.

3.2 The Paradigm of Large-Scale Pre-training

One of the biggest shifts lately is the move to pre-training giant models using huge, loosely matched datasets scraped from the web. Inspired by what worked in NLP, this approach treats multimodal learning as a "pre-train and fine-tune" deal. Models like CLIP (Radford et al., 2021) and ALIGN get trained on hundreds of millions—or even billions—of image-text pairs pulled straight from the internet using a contrastive objective. This pre-training results in a powerful, aligned cross-modal embedding space where semantically similar concepts from different modalities are mapped close together (Jia et al., 2021).

What's wild is that these models don't need a ton of extra training to do new things. You can fine-tune them with just a few examples, or sometimes not at all—they still deliver surprisingly good results. Take CLIP (Radford et al., 2021) , for example. You can show it a prompt like "a photo of a dog," and it'll find dogs in images, even if it never saw labeled dog photos during training. This has opened up powerful vision tools to way more people and pushed innovation in things like open-vocabulary object detection and zero-shot video understanding. Honestly, the sheer scale—both the models themselves and the data they learn from—has turned out to be a big deal for making these systems robust. But there are some real concerns, too. Training these massive models eats up a ton of computing power, and sometimes they pick up on and even amplify social biases baked into huge, messy web datasets. Plus, their black-box nature means it's tough to figure out why they make certain predictions, which can be a headache when you need to trust or debug them in sensitive situations.

4. Conclusion and Future Outlook

This survey takes a deep dive into the fast-moving world of multi-view and multi-modal image representation and generation. We've followed the evolution from handcrafted features to today's deep neural networks and generative AI, pointing out the key architectures, learning tricks, and clever ways of combining information. The rise of big pre-trained models really changed the game, making zero-shot reasoning possible and opening up all sorts of new possibilities for research and real-world use.

Despite the tremendous progress, several challenges remain open for future exploration:

But training and running these giant models—think diffusion models or large vision-language models—demands a ton of computing power. That's a big barrier for most people and not great for the environment, either. So, the next step is to focus on smarter architectures, better training methods, and ways to shrink these models—using neural architecture search, distributed optimization, mixed precision, pruning, quantization, knowledge distillation, and so on—to make them easier for everyone to use.

As these models get more complex and "black-box" like, it gets harder to understand what they're actually doing inside. Being able to control their outputs in a predictable way—like keeping generated videos consistent or tweaking just one part of an image—matters a lot, especially in high-stakes fields like healthcare or self-driving cars. Right now, explainable AI for these multi-modal models is still pretty new and has a long way to go.

Most current models are great at spotting patterns, but they hit a wall with deeper reasoning, figuring out cause and effect, or dealing with structured knowledge. Combining deep learning with symbolic reasoning and knowledge graphs—what people call neuro-symbolic AI—could help models move beyond just statistics and start to show some real understanding, which is what we need for true intelligence.

Looking ahead, the idea of foundation models won't just stick to general-purpose vision-language models. We'll see universal models built specifically for certain types of data or domains—like a foundational model for all kinds of medical imaging, from CT to MRI to X-ray. These models should adapt to different tasks—segmentation, classification, report writing—with very little extra data. Getting there means we need better domain-specific datasets and fresh ideas for model design. And we can't ignore the ethical side: fighting bias in training data, making sure systems are fair and accountable, preventing misuse like deep fakes, and coming up with ways to evaluate all this stuff properly. That's going to take both technical fixes—like bias detection, watermarking, and so on—and smart policy work from the whole community.

So, when you put multi-view learning, multi-modal fusion, and generative AI together, you get machines that can take in way more and actually do something creative with it. The next step is to build systems that aren't just smart, but also efficient, transparent, and reliable—tools that can handle all kinds of data and actually make sense of it. That's how we move closer to real artificial general intelligence. Getting there isn't just about better algorithms or faster computers, though. It's about pushing tech forward while paying attention to how it affects people and society. We're heading from narrow, specialized models to assistants that can do a bit of everything, and innovation on all fronts is what'll get us there.

References

Alemi, A. A., Fischer, I., Dillon, J. V., & Murphy, K. (2016). Deep variational information bottleneck. *arXiv preprint arXiv:1612.00410*. http://arxiv.org/abs/1612.00410

Andrew, G., Arora, R., Bilmes, J., & Livescu, K. (2013). Deep canonical correlation analysis. In *The 30th International Conference on Machine Learning (ICML)* (Vol. 28, pp. 1247–1255).

Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., & Parikh, D. (2015). VQA: Visual question answering. In *The IEEE International Conference on Computer Vision (ICCV)* (pp. 2425–2433). https://doi.org/10.1109/ICCV.2015.279

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In *The 34th International Conference on Machine Learning (ICML)* (Vol. 70, pp. 214–223).

Baltrusaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *The 37th International Conference on Machine Learning (ICML)* (Vol. 119, pp. 1597–1607).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 1, pp. 886–893). https://doi.org/10.1109/CVPR.2005.177

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. http://arxiv.org/abs/2010.11929

Frome, A., Corrado, G. S., Shlens, J., Bengio, S., Dean, J., & Mikolov, T. (2013). DeViSE: A deep visual-semantic embedding model. In *Advances in Neural Information Processing Systems* (Vol. 26, pp. 2121–2129).

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2414–2423).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems* (Vol. 27, pp. 2672–2680).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778).

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 6840–6851).

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika, 28*(3–4), 321–377. https://doi.org/10.2307/2333955

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4700–4708).

Janai, J., Güney, F., Behl, A., & Geiger, A. (2020). Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and Trends in Computer Graphics and Vision, 12*(1–3), 1–308.

Jia, C., Yang, Y., Xia, Y., Chen, Y. T., Parekh, Z., Pham, H., Le, Q. V., Sung, Y., Li, Z., & Duerig, T. (2021). Scaling up visual and vision-language representation learning with noisy text supervision. In *The 38th International Conference on Machine Learning (ICML)* (Vol. 139, pp. 4904–4916).

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4401–4410).

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*. http://arxiv.org/abs/1312.6114

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (Vol. 25, pp. 1097–1105).

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *The IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 10012–10022).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *The 38th International Conference on Machine Learning (ICML)* (Vol. 139, pp. 8748–8763).

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. *arXiv preprint arXiv:2204.06125*. http://arxiv.org/abs/2204.06125

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10684–10695).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. http://arxiv.org/abs/1409.1556

Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. In *Advances in Neural Information Processing Systems* (Vol. 33).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–9).

Tishby, N., Pereira, F. C., & Bialek, W. (2000). The information bottleneck method. *arXiv preprint arXiv:physics/0004057*. https://arxiv.org/abs/physics/0004057

Tsai, Y. H. H., Liang, P. P., Zadeh, A., Morency, L. P., & Salakhutdinov, R. (2019). Learning factorized multimodal representations. In *International Conference on Learning Representations (ICLR)*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998–6008).

Wang, W., Arora, R., Livescu, K., & Bilmes, J. (2015). On deep multi-view representation learning. In *The 32nd International Conference on Machine Learning (ICML)* (Vol. 37, pp. 1083–1092).

Zheng, L., Yang, Y., & Hauptmann, A. G. (2016). Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*. http://arxiv.org/abs/1610.02984

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *The IEEE International Conference on Computer Vision (ICCV)* (pp. 2223–2232).

A RUNWAY-ORIENTED TAKEOFF AND LANDING RISK ANALYSIS USING METAR DATA AND MACHINE LEARNING: A CASE STUDY OF LOEI AIRPORT THAILAND

Peerapong Jantana[1*] and Natdanai Chanlawong[2]

[1,2]System Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand
***Corresponding Author, E-mail:** peeboy0609@gmail.com

Abstract

This study processes METAR data from 2021 to 2025 using a temporal holdout approach in order to reflect real-world time-based operations. Feature engineering was designed with a runway-oriented perspective, including visibility, crosswind, and tailwind components relative to the runway, as well as significant weather phenomena (fog, rain, thunderstorms, and wind gusts) that affect takeoff and landing risks.

The performance of the Decision Tree and Random Forest models was then compared. Reports explicitly marked as NIL were detected and removed to avoid misinterpreting missing data as safe conditions. Data from 2021-2024 were used for model training, while data from 2025 were used for testing.

The results show that the Decision Tree achieved an accuracy of 99.88%, while the Random Forest achieved a higher accuracy of 99.90%, demonstrating superior accuracy and reliability. These findings highlight the potential of integrating aviation meteorological data from METAR reports with machine learning techniques to support aviation safety decision-making under the ICAO Safety Management System (SMS) framework.

Keywords: Machine Learning, METAR, Aviation Weather Hazard, Runway Safety

Introduction

Aviation meteorological services play a critical role in ensuring flight safety, particularly during takeoff and landing phases, which are widely recognized as the most hazardous stages of civil aviation operations. The International Civil Aviation Organization (ICAO) and the World Meteorological Organization (WMO) have established standards and recommended practices for aviation weather observation, reporting, and forecasting to support safety-related decision-making.

Routine aerodrome weather reports (METAR) provide essential real-time meteorological observations at the airport level. Adverse weather conditions such as fog, reduced visibility, strong winds, and thunderstorms significantly increase operational risks during takeoff and landing. Although most airports are equipped with Automated Weather Observation Systems (AWOS), operational risk assessment often relies heavily on the experience of pilots and air traffic controllers, especially under rapidly changing weather conditions.

Accident and incident statistics indicate that a large proportion of aviation safety events occur during takeoff and landing, underscoring the need for systematic and timely decision-support tools capable of assessing weather-related risks. While machine learning techniques have increasingly been applied to aviation meteorology, many existing studies focus on general forecasting or statistical analysis and lack direct linkage to runway-specific operational contexts.

Loei Airport is located in northeastern Thailand and is characterized by basin-like terrain surrounded by mountain ranges, resulting in complex low-level airflow patterns and localized weather phenomena such as morning fog, rapid wind shifts, and frequent thunderstorms during the monsoon season. As a regional airport with limited infrastructure compared to major international airports, effective weather-related risk assessment is essential for maintaining operational safety.

Accordingly, this study proposes a runway-oriented framework for assessing takeoff and landing risks using METAR data and machine learning techniques, aligned with the operational characteristics of Loei Airport and the ICAO Safety Management System (SMS).

Research Objective

1. To analyze takeoff and landing risks by integrating machine learning techniques to support aviation safety decision-making.
2. To analyze historical METAR aviation weather reports in order to identify patterns and trends of adverse weather conditions affecting airport operations and pilots.

Literature Review

Aviation meteorology plays a central role in ensuring operational safety, particularly during takeoff and landing phases, which account for the highest proportion of aviation accidents and incidents worldwide. According to ICAO safety reports, runway excursions, loss of control during approach, and weather-related operational disruptions remain among the most significant safety concerns in civil aviation. Adverse meteorological conditions such as low visibility, precipitation, crosswinds, wind shear, and thunderstorms significantly degrade aircraft performance, increase pilot workload, and reduce decision margins during critical flight phases.

METAR (Meteorological Aerodrome Report) constitutes the primary standardized source of surface aviation weather observations. It provides real-time aerodrome-level data including wind, visibility, cloud conditions, weather phenomena, temperature, and pressure. Several studies have demonstrated that METAR data can be effectively utilized beyond operational reporting purposes and applied in predictive analytics and safety assessment frameworks.

Recent advancements in machine learning have enabled more sophisticated modeling of aviation weather hazards. Kharisma et al. (2025) applied Random Forest regression to predict airport visibility using integrated METAR parameters, demonstrating improved short-term prediction accuracy under low-visibility conditions. Their results confirm that ensemble learning methods can effectively capture nonlinear relationships among meteorological variables.

Similarly, Muangsong (2023) employed machine learning techniques for thunderstorm nowcasting based solely on METAR observations at northern Thailand airports. The study demonstrated that short-term hazard prediction can be achieved using routine aerodrome reports, supporting operational decision-making without relying exclusively on numerical weather prediction models.

Beyond forecasting applications, machine learning has also been used for hazardous weather classification. Patriarca et al. (2022) proposed anomaly detection and hierarchical clustering techniques to support performance management in aerodrome weather forecasting systems. Their findings emphasize the importance of interpretable models for operational environments where transparency is critical.

In runway safety research, Zhu et al. (2024) developed a real-time landing risk assessment framework integrating finite element modeling and machine learning under wet runway conditions. Their work highlights the limitations of traditional deterministic simulation methods and underscores the need for data-driven, real-time risk assessment tools. Wet runway conditions combined with precipitation significantly increase braking distance and the probability of runway excursion events.

Weather-related accident analyses further reinforce this concern. Kyaw Than Oo et al. (2023) reported that meteorological hazards contributed substantially to aviation accidents and delays at Yangon Airport, particularly during monsoon seasons. Their climatological assessment demonstrated that visibility degradation and convective weather phenomena were dominant risk factors.

More recently, CHOMA et al. (2025) explored artificial intelligence techniques for hazardous weather prediction in aviation environments. Their work supports the adoption of ensemble learning and decision-tree-based models due to their balance between predictive performance and interpretability—an essential characteristic for safety-critical domains.

Additionally, Gdeeb (2024) proposed a meta-based Random Forest fusion framework for weather classification, showing that ensemble approaches consistently outperform single-model classifiers in complex atmospheric conditions. These findings are consistent with broader machine learning literature, where Random Forest models demonstrate robustness against overfitting and strong generalization capability.

Although previous studies have successfully applied machine learning to aviation meteorology, most research has primarily focused on general weather forecasting, visibility prediction, thunderstorm detection, and flight delay analysis.

However, relatively limited attention has been given to runway-oriented operational risk assessment, particularly approaches that integrate runway heading, crosswind and tailwind components, aviation safety thresholds, and ICAO Safety Management System (SMS) principles.

Traditional airport risk evaluation typically relies on fixed meteorological thresholds or operator experience. While effective to some extent, such approaches fail to fully capture the complex and nonlinear interactions among multiple meteorological variables. The ICAO SMS framework advocates a proactive, data-driven, and risk-based approach to hazard identification and mitigation. Therefore, integrating machine learning with runway-specific meteorological feature engineering represents a logical advancement toward predictive aviation safety management.

Ensemble models such as Random Forest further offer the advantage of interpretability through feature importance analysis. In safety-critical environments, model transparency is essential to ensure stakeholder trust, including pilots, air traffic controllers, and safety managers.

Despite the growing body of research in aviation weather analytics, several critical gaps remain:

(1) runway-specific hazard modeling,
(2) regional airport case studies in Southeast Asia,
(3) integration of METAR-coded weather phenomena with quantitative variables, and
(4) temporal holdout validation that simulates real-world forecasting conditions.

This study addresses these gaps by proposing a runway-oriented machine learning framework that incorporates crosswind and tailwind calculations relative to Runway 01 at Loei Airport. By aligning feature engineering with operational aviation logic and SMS risk principles, the research advances beyond general weather prediction toward structured, operational risk assessment in real-world contexts.

Research Methodology

1. Data Preparation

This study utilizes aviation meteorological data (METAR) from Loei Airport covering the period from 2021 to 2025. The METAR data were retrieved from the Ogimet database. Hourly observations were collected and stored in .txt format, resulting in 43,573 records. Reports marked as NIL were removed prior to analysis to ensure data completeness, leaving 26,521 valid records.

The dataset includes various meteorological parameters such as wind speed (knots), visibility (kilometers), weather phenomena including thunderstorms (TSRA/TS), rain, mist, fog, cloud base height (feet), temperature (°C), and atmospheric pressure (millibars).

2. Feature Engineering

Runway-oriented wind components were calculated based on the actual runway configuration, specifically Runway 01 (heading 010°). Crosswind and tailwind components were derived using trigonometric equations to assess their direct impact on aircraft controllability. Additional hazardous weather indicators, including fog, rain, thunderstorms, and wind gusts, were incorporated into the feature set.

$$\text{Crosswind} = V \sin \theta \quad (1)$$

$$\text{Tailwind} = V \cos \theta \quad (2)$$

3. Hazard Identification

The identification of "hazardous" conditions in this study is not based solely on general statistical thresholds, but rather on aviation safety logic, which reflects the risk management principles used by pilots and air traffic controllers. A hazard is assigned a value of 1 when any one of the conditions listed in Table 1 is met.

Table 1: Hazard Decision Criteria

Parameters	Condition
Thunderstorm	TSRA/TS detected
Rain	+RA/RA detected
Mist	BR detected, 1,000 m. ≤ Visibility < 5,000 m.
Fog	FG detected, Visibility < 1,000 m.
Low Visibility	Visibility ≤ 2,000 m.
Wind Speed	Wind ≥ 15 kt
Gust	Wind ≥ 20 kt
Crosswind RWY01	Wind ≥ 15 kt
Tailwind RWY01	Wind ≥ 10 kt

4. Machine Learning Models and Temporal Validation

This study applies machine learning techniques to develop models for analyzing takeoff and landing risks using aviation meteorological data from METAR reports. The selected algorithms are Decision Tree and Random Forest, as they provide interpretable logical structures and effectively handle nonlinear data.

The Decision Tree model uses a tree-based structure to split data based on feature conditions. The tree depth was limited to 6 levels (max depth = 6), with a minimum of 10 samples per node (min samples = 10) and a fixed random state (random state = 42) to reduce noise and improve decision stability.

The Random Forest model is an ensemble learning technique consisting of multiple decision trees trained on different subsets of data and features. Predictions from individual trees are aggregated using majority voting, reducing overfitting and improving predictive accuracy compared to a single decision tree. The Random Forest model was configured with 200 trees, a minimum of 10 samples per node, and random state = 42.

To reflect real-world forecasting scenarios, where pilots and ATC rely on past data to anticipate future conditions, a time-series-based data split was applied instead of random sampling. Data from 2021–2024 (21,626 samples) were used for training, while data from 2025 (4,895 samples) were used for testing, simulating real operational conditions. The overall research workflow is illustrated in Figure 1.

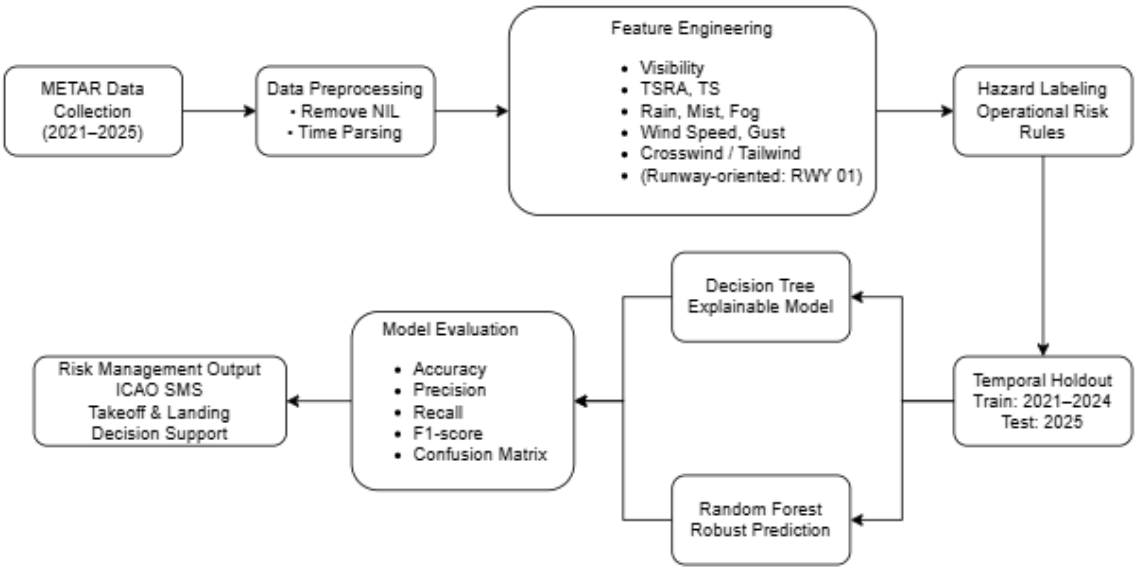


Figure 1: Research Workflow Diagram

Results

This section presents the performance of the proposed machine learning models for assessing weather-related takeoff and landing risks at Loei Airport (VTUL). The analysis focuses on Runway 01 and employs a temporal holdout strategy to ensure that the evaluation reflects realistic operational conditions.

The Decision Tree and Random Forest models were trained using METAR data from 2021 to 2024 and evaluated using data from 2025. This approach simulates real-world deployment scenarios, in which models must predict future operational risks based solely on past observations, while also preventing information leakage. Overall, both models demonstrated excellent predictive performance. The Decision Tree achieved an accuracy of 99.88%, while the Random Forest achieved a slightly higher accuracy of 99.90%. Although the difference in accuracy is marginal, the Random Forest model consistently exhibited superior performance across all evaluation metrics, particularly in terms of Recall and F1-score. Comparative results are summarized in Table 2.

Table 2: Compares the performance of the models

	Decision Tree	Random Forest
Accuracy	99.88%	99.90%
Precision	99.71%	99.72%
Recall	99.59%	99.87%
F1-score	99.15%	99.29%

From an aviation safety perspective, Recall is a critical indicator because it reflects the model's ability to correctly identify hazardous operational conditions. The Random Forest model achieved a higher Recall value, indicating a lower likelihood of failing to detect weather conditions that pose risks during takeoff and landing.

Further insight is provided by the confusion matrix analysis. The majority of test samples were correctly classified as either hazardous or non-hazardous conditions. Misclassifications occurred only in a small number of cases. Importantly, the Random Forest model produced fewer false negative predictions than the Decision Tree model. This is a key advantage in safety-critical applications, as false negatives represent situations in which hazardous conditions are incorrectly classified as safe (Figure 2). False positive predictions, where non-hazardous conditions were classified as hazardous, were observed at a low rate in both models. While such errors may lead to conservative operational decisions, they are generally less critical than false negatives within the context of aviation safety management.

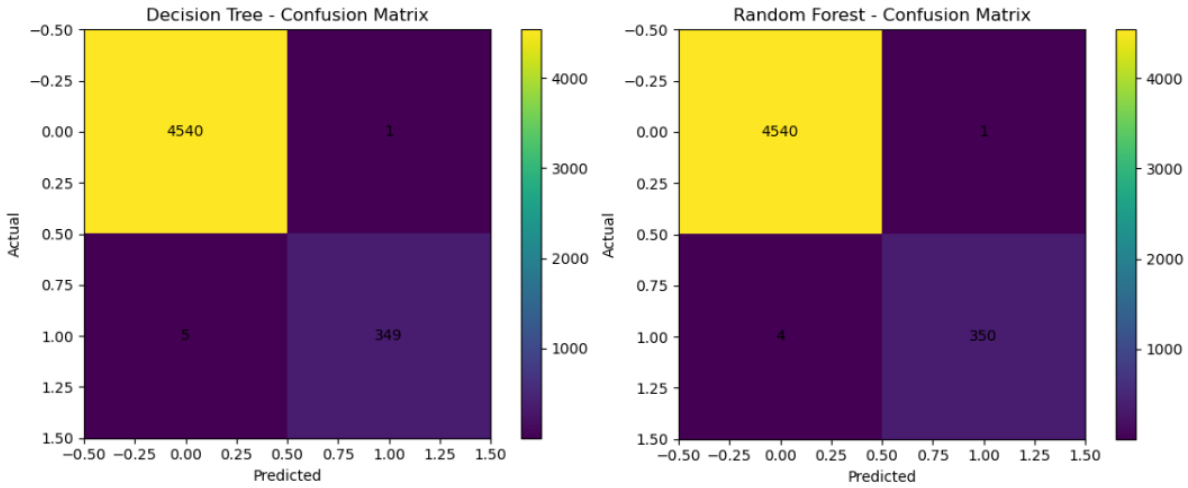


Figure 2: Confusion Matrix Decision Tree and Random Forest

An analysis of feature importance derived from the Random Forest model reveals that precipitation-related phenomena, particularly rain (RA), exert the strongest influence on risk classification. Mist (BR) and numerical visibility values were also identified as highly influential variables, underscoring the dominant role of visibility degradation in takeoff and landing safety. These results are consistent with established aviation meteorology principles, which recognize reduced visibility and wet runway conditions as major contributors to operational risk. In contrast, wind-related variables such as crosswind, tailwind, wind speed, and gusts exhibited lower relative importance. This does not diminish their operational relevance; rather, it suggests that within the climatological and operational context of Loei Airport, severe wind conditions occur less frequently or are effectively mitigated by existing operational procedures (Figure 3).

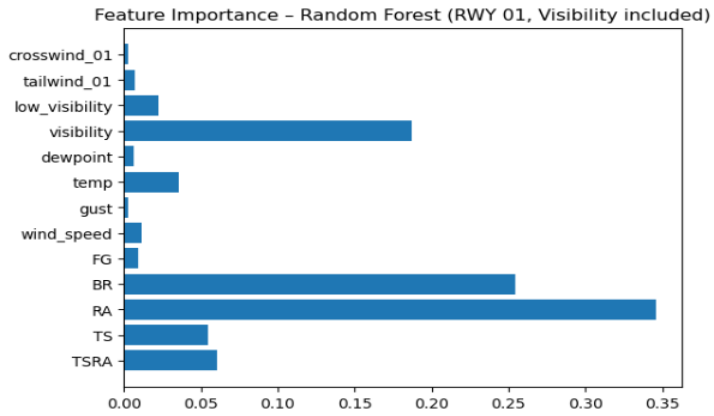


Figure 3: Feature Importance Analysis

In summary, the results demonstrate that the proposed runway-oriented machine learning framework can reliably identify weather-related hazards affecting takeoff and landing operations. The Random Forest model, in particular, provides a robust balance between predictive accuracy and operational reliability, making it well-suited for integration into aviation weather risk assessment and decision-support systems.

Discussion and Conclusion

This study proposed a runway-oriented framework for assessing aircraft takeoff and landing risks using METAR data integrated with machine learning techniques, with Loei Airport (VTUL) as a case study. By explicitly incorporating runway direction, wind components, and operationally relevant weather phenomena, the proposed approach bridges the gap between meteorological observations and practical aviation safety decision-making.

The experimental results demonstrate that both the Decision Tree and Random Forest models are highly effective in classifying hazardous and non-hazardous operational conditions. The Random Forest model consistently outperformed the Decision Tree, achieving higher overall accuracy and, more importantly, reducing the number of false negative predictions. This characteristic is particularly critical in aviation safety applications, where failure to detect hazardous conditions may lead to severe operational consequences.

Analysis of feature importance revealed that precipitation-related phenomena, especially rain (RA), along with mist (BR) and reduced visibility, were the most influential factors contributing to takeoff and landing risk at Loei Airport. These findings are consistent with established principles of aviation meteorology, which emphasize the impact of visibility degradation and runway surface contamination on aircraft controllability, braking performance, and pilot workload during approach and landing phases. Notably, qualitative METAR weather codes were found to provide stronger predictive influence than numerical visibility values alone, highlighting the importance of contextual meteorological interpretation rather than reliance on fixed numerical thresholds.

Although wind-related hazards such as crosswind, tailwind, and gusts are generally regarded as critical risk factors in aviation operations, their statistical importance in this study was relatively lower. This may be attributed to the lower frequency of extreme wind events at the study airport or to effective operational mitigation measures already in place. Nevertheless, wind-related variables remain operationally significant and should continue to be considered in comprehensive risk assessment frameworks.

Overall, the results confirm the suitability of Random Forest as a robust and interpretable machine learning model for aviation weather risk assessment. The ability to explain model behavior through feature importance analysis aligns well with the principles of the ICAO Safety Management System (SMS), which emphasizes transparency, risk-based decision-making, and proactive hazard identification.

In conclusion, this study demonstrates that the integration of METAR data, runway-oriented feature engineering, and machine learning techniques can provide an effective and systematic approach for assessing weather-related takeoff and landing risks. The proposed framework offers a practical foundation for the development of weather-based decision support tools aimed at enhancing operational safety at regional airports. Future research should extend this approach by incorporating additional operational variables, real-time weather data, and multi-airport validation to further improve generalizability and support proactive aviation safety management.

Recommendations

Future research may expand this framework by incorporating additional operational data such as runway surface condition reports, aircraft type, and traffic volume. Integrating real-time weather observations and forecasts, including TAF and nowcasting systems, would enable a transition from reactive to predictive risk management in line with ICAO's proactive safety philosophy. Further validation across different airports and runway configurations is recommended to assess the generalizability of the proposed approach.

References

Choma, L., Kelemen, M., Jr., Antosko, M., Ozdincova, K., & Sabo, J. (2025). Prediction of hazardous weather phenomena using artificial intelligence. *Acta Avionica*, *27*(52). https://doi.org/10.35116/aa.2025.0006

Gdeeb, S. T. (2024). Weather classification using meta-based random forest fusion of transfer learning models. *International Journal of Advanced Intelligent Informatics*, *10*(2), 186-201. https://doi.org/10.26555/ijain.v10i2.1264

Kharisma, A., Fadhillah, M., & Haryanto, Y. D. (2025). Advancing aviation meteorology: Airport visibility prediction using random forest regressor on integrated METAR parameters. *Jurnal Ilmu dan Inovasi Fisika,9(2),* 62-72. https://doi.org/10.24198/jiif.v9i2.65464

Muangsong, C. (2023). Thunderstorm nowcasting from meteorological terminal air reports using machine learning techniques: A case study of airports in northern Thailand. https://hdl.handle.net/20.500.14770/13834

Oo, K. T., Jonah, K., & Oo, K. L. (2023). A systematic climatology report of aviation weather hazard on Yangon airport region. *Journal of Meteorological Research and Applications, 1*(2), 89-103. https://doi.org/10.3126/jomra.v1i2.61191

Patriarca, R., Simone, F., & Di Gravio, G. (2022). Supporting weather forecasting performance management at aerodromes through anomaly detection and hierarchical clustering. *Expert Systems with Applications*, *213*, 119210. https://doi.org/10.1016/j.eswa.2022.119210

World Meteorological Organization. (2021). *Guide to aeronautical meteorological services* (WMO-No. 732).

Zhu, X., Wu, Y., Yang, Y., Pang, Y., Ling, H., & Zhang, D. (2024). Real-time risk assessment of aircraft landing based on finite element-virtual prototype-machine learning co-simulation on wet runways. *International Journal of Transportation Science and Technology*, *13*, 77-90. https://doi.org/10.1016/j.ijtst.2023.11.007

HIGH-RESOLUTION ROAD SURFACE TEMPERATURE MAPPING USING UAV DATA AND LANDSAT THERMAL OBSERVATIONS

Chayanon Penyalai[1], Kittipong Prapakan[2], Jurawan Nontapon[3], and Siwa Kaewplang[4*]

[1,2,4]Faculty of Engineering, Mahasarakham University, Mahasarakham, Thailand
[3]Faculty of Science, Chandrakasem Rajabhat University, Bangkok, Thailand
[*]**Corresponding Author, E-mail**: siwa.kae@msu.ac.th

Abstract

Rapid urban expansion intensifies the Urban Heat Island (UHI) effect, with asphalt road surfaces acting as major thermal hotspots. However, satellite-derived Land Surface Temperature (LST) products are limited by coarse spatial resolution, while UAV-based observations lack temporal continuity. This study proposes a fusion-driven thermal downscaling framework that integrates high-resolution UAV RGB imagery (0.20 m) with Landsat 8–9 thermal observations to generate spatially refined Road Surface Temperature (RST) maps. The framework combines thermal sharpening, bias correction, and machine learning regression using Random Forest (RF) and Support Vector Regression (SVR). Model performance was evaluated using five-fold cross-validation.

Results indicate that the RF model significantly outperformed SVR, achieving a testing R^2 of 0.66 with an RMSE of 1.99°C, demonstrating reliable generalization under heterogeneous urban surface conditions. Feature importance analysis revealed that bias-corrected Landsat LST was the dominant predictor, while UAV-derived spectral indices refined sub-pixel thermal variability along the road corridor. The proposed fusion framework preserves the physical thermal support of satellite observations while enhancing spatial detail to 0.20 m resolution, providing a statistically defensible approach for high-resolution urban pavement thermal assessment.

This approach offers practical potential for urban heat mitigation planning and heat-resilient infrastructure design in rapidly warming cities.

Keywords: Urban Heat Island (UHI), Road Surface Temperature (RST), Thermal Downscaling, UAV–Satellite Data Fusion, Random Forest Regression

Introduction

The intensification of the Urban Heat Island (UHI) phenomenon is a direct consequence of rapid urban expansion, where anthropogenic surfaces—particularly asphalt and concrete pavements—exhibit high heat-retention capacity and low reflectance, resulting in elevated Road Surface Temperature (RST) compared to natural land covers (Dai et al., 2025; Wu et al., 2022). Road infrastructures account for a large proportion of impervious urban surfaces and are recognized as major thermal hotspots due to continuous solar radiation absorption and anthropogenic heat release from vehicular activity (Kawakubo et al., 2023). Recent remote sensing investigations indicate that micro-scale variability in RST is influenced not only by road material type but also by urban form, surface roughness, and surrounding land cover configuration (Andriambololonaharisoamalala et al., 2025; Henn et al., 2024). Satellite-

derived Land Surface Temperature (LST) products, such as those from Landsat and MODIS, provide consistent temporal monitoring capabilities but are constrained by coarse spatial resolution, limiting their effectiveness in characterizing fine-scale thermal heterogeneity along road corridors (Bahi et al., 2025; Aboutalebi et al., 2022). In contrast, Unmanned Aerial Vehicle (UAV) platforms enable centimeter-level thermal assessments but lack temporal coverage and are subject to operational constraints such as flight regulations and atmospheric conditions (Dai et al., 2025; Sun et al., 2019). This creates a spatial–temporal resolution gap in current UHI-oriented thermal monitoring frameworks. To overcome these limitations, emerging research has focused on multi-sensor fusion and thermal downscaling techniques, integrating satellite thermal data with high-resolution UAV spectral information to enhance urban microclimate modeling accuracy (Wu et al., 2022; Su et al., 2025). Studies by Dai et al. (2025) and Wu et al. (2022) demonstrated the capability of machine learning-based downscaling frameworks to preserve thermal gradients across heterogeneous surfaces, while Aboutalebi et al. (2022) and Su et al. (2025) reported substantial improvements in temperature prediction accuracy through UAV–satellite data fusion. However, the majority of these studies emphasize vegetation or agricultural thermal monitoring, with significantly fewer addressing the thermal behavior of urban road infrastructure, which presents unique radiometric characteristics due to material composition and dynamic anthropogenic influence (Kawakubo et al., 2023; Henn et al., 2024; Zargari et al., 2024). Therefore, this study introduces a fusion-driven modeling framework that integrates UAV RGB spectral indices with Landsat-derived LST using machine learning regression to generate high-resolution RST maps. By leveraging the temporal consistency of satellite thermal imaging and the spatial precision of UAV-based sensing, this research aims to advance thermographic diagnostics for urban pavement systems and provide a data-driven foundation for heat-resilient infrastructure planning in the context of increasing UHI intensity.

Research Objectives

1. To develop a multi-sensor data fusion framework integrating UAV RGB imagery and Landsat thermal data to generate sub-meter resolution road surface temperature maps.
2. To evaluate machine learning models for urban thermal prediction by comparing Random Forest, Support Vector Regression, and Decision Tree algorithms in estimating pavement temperature under heterogeneous urban conditions.
3. To assess the applicability of high-resolution thermal maps for smart city heat mitigation planning**,** supporting heat-resilient infrastructure design and urban heat island management strategies.

Literature Review

1. Urban Heat Island and the Thermal Behavior of Road Surfaces

The Urban Heat Island (UHI) phenomenon is strongly associated with the replacement of natural land cover by impervious urban materials such as asphalt and concrete pavements. These materials typically exhibit low albedo and high thermal inertia, enabling them to store solar energy and release heat slowly, thereby forming persistent thermal hotspots within cities. Road networks therefore play a critical role in localized heat accumulation and contribute significantly to urban microclimate alteration (Kawakubo et al., 2023; Zargari et al., 2024).

Previous studies have shown that Road Surface Temperature (RST) variability depends not only on pavement material but also on urban morphology, surface roughness, and surrounding land cover configuration (Dai et al., 2025; Sun et al., 2019). Consequently, fine-scale thermal observation is required to accurately characterize thermal heterogeneity along road corridors.

However, many previous studies have focused primarily on general urban land surface temperature patterns rather than specifically examining road infrastructure at fine spatial scales. As a result, the thermal behavior of road surfaces within complex urban environments remains insufficiently explored.

2. Satellite-Derived Land Surface Temperature and Its Limitations

Satellite-based Land Surface Temperature (LST) products derived from sensors such as Landsat and MODIS provide consistent temporal monitoring and reliable radiometric thermal measurements, making them widely used in UHI studies (Wu et al., 2022; Andriambololonaharisoamalala et al., 2025). Landsat thermal sensors typically provide spatial resolutions of approximately 100 m (resampled to 30 m), which remain relatively coarse compared to the spatial scale of urban infrastructure.

Because each pixel frequently contains mixed land covers such as pavement, vegetation, and soil, satellite-derived LST often suffers from spatial smoothing and loss of micro-scale temperature variability (Bahi et al., 2025). This limitation restricts its ability to accurately capture localized thermal anomalies along road networks.

Although several thermal downscaling techniques have been proposed to improve spatial detail (Aboutalebi et al., 2022; Su et al., 2025), many approaches still rely primarily on satellite-derived predictors. Consequently, the integration of higher-resolution spatial information remains limited in many urban temperature studies.

3. UAV Remote Sensing for High-Resolution Urban Surface Analysis

Unmanned Aerial Vehicle (UAV) imagery provides centimeter-level spatial resolution, allowing precise identification of pavement boundaries, roadside vegetation, and built structures that influence local thermal patterns (Dai et al., 2025; Henn et al., 2024). Such spatial detail is particularly valuable for analyzing temperature heterogeneity within urban environments.

However, UAV observations typically lack temporal continuity and stable radiometric calibration, especially when only RGB sensors are used. While UAV-mounted thermal cameras can capture instantaneous temperature patterns, their use is constrained by operational conditions and limited acquisition frequency (Henn et al., 2024).

Therefore, UAV and satellite observations possess complementary characteristics: satellite data provide reliable thermal signals and temporal consistency, whereas UAV imagery offers highly detailed spatial information. Nevertheless, the integration of UAV-derived spatial predictors with satellite thermal observations remains relatively limited in existing research.

4. Multi-Sensor Data Fusion and Thermal Downscaling

To overcome the spatial–temporal trade-off in remote sensing observations, recent research has explored multi-sensor data fusion techniques that integrate satellite thermal data with high-resolution spatial predictors (Wu et al., 2022; Bahi et al., 2025). Thermal downscaling frameworks attempt to reconstruct fine-scale temperature patterns by modeling the relationships between coarse-resolution thermal observations and high-resolution auxiliary variables.

Machine learning-based downscaling approaches have demonstrated strong capability in modeling nonlinear relationships between spectral variables and temperature (Aboutalebi et al., 2022; Su et al., 2025). By combining macro-scale thermal information from satellites with micro-scale spatial variability from high-resolution imagery, these fusion frameworks significantly improve temperature estimation accuracy (Andriambololonaharisoamalala et al., 2025).

However, most existing downscaling studies primarily focus on vegetation and agricultural applications. Urban road infrastructure, which possesses unique radiometric properties and strong anthropogenic heat influences, has received comparatively less attention in thermal downscaling research.

5. Machine Learning Models for Surface Temperature Prediction

Machine learning approaches are widely adopted for surface temperature estimation because they can effectively capture nonlinear relationships and interactions among multiple predictors. Random Forest (RF) models are particularly popular in remote sensing applications due to their ensemble learning structure and strong predictive robustness (Wang et al., 2019; Wei et al., 2020).

Support Vector Regression (SVR) has also been widely used for temperature prediction because of its ability to model complex relationships with relatively limited input features. However, SVR performance is sensitive to parameter tuning and kernel selection (Su et al., 2025). Classification and Regression Tree (CART) models offer interpretability and simple decision rules, but they often exhibit lower stability and predictive accuracy when applied to complex environmental datasets (Zargari et al., 2024).

Comparative studies consistently indicate that ensemble learning approaches such as Random Forest generally provide superior predictive performance and generalization capability for urban thermal prediction tasks (Wang et al., 2019).

6. Feature Selection and Spectral Redundancy

High-resolution imagery allows the extraction of numerous spectral indices and spatial features that describe surface reflectance, material composition, and urban structure. However, many of these predictors are highly correlated, which can introduce spectral redundancy and increase model complexity (Dai et al., 2025).

Feature selection techniques such as permutation importance and cross-validation-based reduction are therefore essential for identifying the most relevant predictors and preventing model overfitting (Wang et al., 2019; Wei et al., 2020). Reducing redundant variables not only improves model interpretability but also enhances predictive stability in multi-sensor thermal fusion frameworks.

Despite these methodological advances, limited studies have investigated feature selection strategies specifically for road surface temperature prediction using UAV-derived predictors combined with satellite thermal observations.

Research Methodology

A. Study Area

The study was conducted within the premises of Mahasarakham University (MSU), located along Highway No. 2202 in Kham Riang Subdistrict, Kantharawichai District, Mahasarakham Province, Thailand

The central coordinate of the study site is approximately 16.244515°N, 103.249563°E. The selected study site is an asphalt road approximately 1 km in length, consisting of four traffic lanes, each 3 m wide. The surrounding area includes residential zones, agricultural fields, and small green patches, representing a heterogeneous pattern of urban land use within Mahasarakham City.

Mahasarakham Province lies in a tropical monsoonal climate zone, characterized by persistently warm conditions throughout the year. The annual mean temperature ranges between 17°C and 35°C, with extremes rarely falling below 13°C or exceeding 39°C. The province experiences a

prolonged wet season lasting approximately 8.9 months (February–November) and a dry season of about 3.1 months (November–February). The annual average precipitation is around 1,200 mm, with August being the wettest month (average rainfall ≈ 214 mm) and December the driest month (≈ 2 mm).

Overall, the region exhibits a distinct monsoonal pattern with alternating wet and dry periods, making it an ideal setting for examining surface temperature variability and thermal behavior of urban road surfaces under tropical environmental conditions. The location of the study area is show in Figure 1.

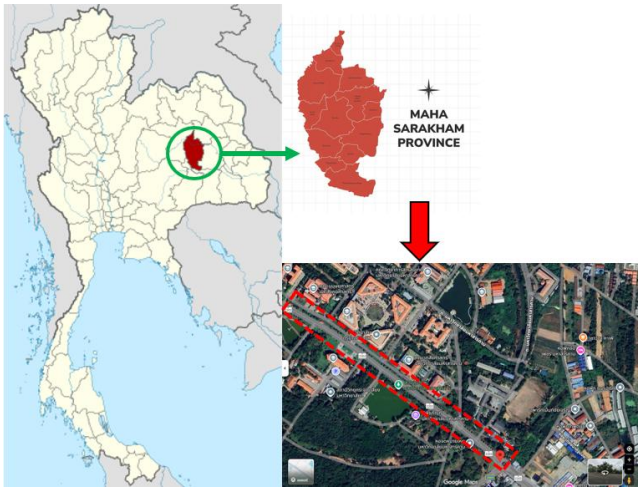


Figure 1:. Location of the study area and road in mahasarakham university.

B. Field Data Collection

Field measurements of road surface temperature (Temp) were conducted using a calibrated digital thermal gun positioned approximately 1.0 m above the surface. The geographic coordinates of each observation were obtained with an RTK GNSS receiver, achieving a horizontal accuracy of ±4 cm. A total of 44 sampling locations were established, spaced at intervals exceeding 100 m to correspond with the 30 m spatial resolution of the Landsat-derived Land Surface Temperature (LST) data.

The survey was carried out on 8 June 2025, between 11:30 and 14:30 local time, coinciding with the Landsat overpass window to ensure temporal consistency between field observations and satellite acquisitions. This period represents the peak daily heating phase, capturing the maximum variability of surface temperature conditions across the study area.

C. Satellite and UAV Data

This study integrates high-resolution UAV RGB imagery with Landsat 8–9 thermal data using a physically consistent, model-based thermal downscaling framework. The objective was to preserve the original thermal support of Landsat while exploiting the fine spatial detail of UAV imagery to model sub-pixel temperature variability.

The UAV dataset consisted of an orthorectified RGB mosaic with a ground sampling distance (GSD) of 0.20 m. The orthomosaic was geometrically corrected and co-registered to the Landsat reference coordinate system (WGS84/UTM Zone 48N). From the UAV RGB bands (Red, Green, Blue), several spectral indices were computed to characterize surface reflectance variability and material differences influencing thermal behavior. These included GRVI, VARI, Excess Green (ExG), and Brightness, which are commonly used to capture vegetation condition, surface color contrast, and radiometric variability.

Landsat 8 and 9 Collection 2 Level-2 products were used to derive Land Surface Temperature (LST) from TIRS Band 10. Thermal values were converted from scaled digital numbers to Kelvin and subsequently to degrees Celsius using standard radiometric calibration coefficients. Cloud and cloud-shadow contamination were removed using QA_PIXEL bitmask filtering. The resulting LST image was composited (median) over the study period and maintained at its native 100 m spatial resolution to preserve physical thermal support.

Field-measured near-surface temperature (Temp Road) points were spatially linked to the aggregated dataset for model validation and interpretation. All predictors were standardized prior to model development. This preprocessing workflow ensured geometric alignment, spectral consistency, and physical validity of thermal information prior to statistical modeling.

D. Model Development

This study developed a fusion-based thermal downscaling framework to estimate high-resolution Road Surface Temperature (RST) by integrating Landsat-derived thermal information with high-resolution UAV RGB predictors. The overall procedure consisted of four main steps: (1) preparation of the coarse-resolution thermal target, (2) aggregation of UAV predictors to match the thermal support, (3) thermal sharpening to distribute temperature variability within each Landsat pixel, and (4) bias correction to preserve the original satellite-derived thermal signal.

1) Preparation of the Landsat thermal target

Landsat 8 and Landsat 9 Collection 2 Level-2 images acquired during May–July 2025 were used to derive Land Surface Temperature (LST) from Thermal Infrared Sensor (TIRS) Band 10. Cloud and cloud-shadow pixels were removed using the QA_PIXEL band. The thermal band was converted from scaled digital numbers to temperature values in degrees Celsius following the standard radiometric scaling procedure provided for Collection 2 Level-2 products. To reduce noise associated with individual scenes and short-term atmospheric variability, a median thermal composite was generated for the study period. The resulting LST image was maintained at its native thermal support of 100 m, which was treated as the physically meaningful reference scale for model development.

2) Aggregation of UAV predictors to thermal support

The UAV orthomosaic (0.20 m spatial resolution) was used to derive a set of fine-scale predictors, including the RGB bands and spectral indices such as GRVI, VARI, Excess Green (ExG), and Brightness. Because Landsat LST represents the average thermal condition of a much larger ground area, these UAV-derived variables were aggregated to the same 100 m grid as the Landsat thermal pixels. For each Landsat pixel, summary statistics of the UAV predictors were calculated so that the explanatory variables and the thermal response variable were spatially consistent. This step is important because it prevents scale mismatch and ensures that the machine learning models are trained using predictors that correspond to the same physical support as the observed thermal signal.

3) Thermal sharpening

Thermal sharpening was then performed to infer sub-pixel thermal variability within each 100 m Landsat pixel. A regression relationship was first established between the aggregated UAV predictors

and the Landsat-derived LST at the 100 m scale. In this study, Random Forest (RF) and Support Vector Regression (SVR) models were trained to learn the nonlinear relationship between surface reflectance characteristics and thermal conditions. After model calibration, the trained regression model was applied back to the original 0.20 m UAV predictor layers. This generated an initial high-resolution temperature surface, here referred to as the sharpened temperature map.

The sharpened map represents the spatial distribution of temperature variation within each coarse thermal pixel, guided by the fine-scale surface characteristics captured by UAV imagery. In other words, the model uses local differences in road color, brightness, vegetation influence, and surface texture proxies to redistribute the coarse Landsat temperature into a more spatially detailed pattern. However, because this prediction is obtained statistically, the average temperature of the sharpened pixels within a Landsat block may not exactly equal the original Landsat LST value. Therefore, an additional bias correction step was required.

4) Bias correction

To preserve physical consistency with the original satellite thermal observations, a block-wise bias correction was applied to the sharpened temperature map. For each 100 m Landsat pixel, the mean value of all sharpened 0.20 m predictions located inside that pixel was calculated. The difference between this mean sharpened temperature and the original Landsat LST was then computed as a residual bias term:

$$Bias_i = LST_i - \overline{LST_{sharp,i}}$$

where LST_i is the original Landsat temperature of pixel (i), and $\overline{LST_{sharp,i}}$ is the mean of the sharpened predictions within that same pixel.

This residual bias was added uniformly to all high-resolution sharpened pixels inside the corresponding 100 m block: $LST_{sharp_{bc},ij} = LST_{sharp,ij} + Bias_i$

where $LST_{sharp,ij}$ is the initial sharpened prediction at sub-pixel location (j) within Landsat pixel (i), and $LST_{sharp_{bc},ij}$ is the bias-corrected sharpened temperature.

This procedure ensures that the average of the corrected high-resolution temperatures within each Landsat pixel exactly matches the original Landsat-derived LST. Therefore, the bias correction preserves thermal energy consistency at the coarse scale while retaining the fine-scale spatial pattern introduced during sharpening. As a result, the final temperature map is both spatially refined and physically constrained by the satellite observation.

5) Road surface temperature prediction and validation

The bias-corrected sharpened LST layer $LST_{sharp_{bc}}$ was then integrated with UAV-derived predictors and field-observed road surface temperature samples (Temp Road) to develop the final road temperature prediction models. RF and SVR were trained and evaluated using five-fold cross-validation with pooled out-of-fold predictions. Model performance was assessed using the coefficient of determination (R^2)

and Root Mean Square Error (RMSE). In addition, wrapper-based variable selection and hyperparameter tuning were performed under a stability criterion of $|Train\ R^2 - Validation\ R^2| \leq 0.10$ to reduce overfitting and improve model generalization.

Importantly, this framework does not create independent thermal observations at 0.20 m resolution. Instead, it statistically refines the spatial distribution of temperature within each Landsat thermal pixel while preserving the original thermal support of the satellite data. Thus, the final RST product should be interpreted as a physically constrained, high-resolution estimation of sub-pixel thermal variability rather than as direct 0.20 m thermal measurement.

Results

1. Model Performance Evaluation

The predictive performance of the machine learning models is summarized in Table 1, while the relationships between observed and predicted road surface temperatures are illustrated in Figure 2:

As shown in Table 1, the Random Forest (RF) model demonstrated superior predictive capability compared to the Support Vector Regression (SVR) model. The RF model achieved a training R^2 of 0.75 with an RMSE of 1.18°C, and a testing R^2 of 0.66 with an RMSE of 1.99°C. The relatively small difference between training and testing R^2 values ($\Delta R^2 = 0.09$) indicates good generalization ability and limited overfitting under the five-fold cross-validation framework.

In contrast, the SVR model exhibited considerably lower performance, with training and testing R^2 values of 0.39 and 0.31, respectively, and RMSE values exceeding 1.95°C. The modest difference between training and testing accuracy suggests stable but weak predictive capability, indicating that SVR was less effective in capturing the nonlinear relationships governing road surface thermal variability.

The scatter plots in Figure 2: further illustrate these differences. In Figure. 2(a), RF predictions closely align with the 1:1 reference line for both training (green) and testing (red) datasets, particularly within the dominant temperature range (approximately 51–55°C). Although slight dispersion is observed at lower temperature values, the overall distribution indicates strong agreement between observed and predicted temperatures.

Conversely, Figure. 2(b) shows that SVR predictions display greater deviation from the 1:1 line and increased scatter across the full temperature range. The model exhibits reduced sensitivity to higher temperature values, resulting in systematic underestimation at the upper range and weaker correlation overall.

These results confirm that the RF model provides more reliable and accurate road surface temperature estimation under heterogeneous urban surface conditions.

Table 1: Comparative performance of machine learning models (RF, SVR, and CART) based on R^2 and RMSE for training and testing datasets.

Model	Train		Validation	
	R square	RMSE	R square	RMSE
RF	0.75	1.18	0.66	1.99
SVR	0.39	1.95	0.31	2.01

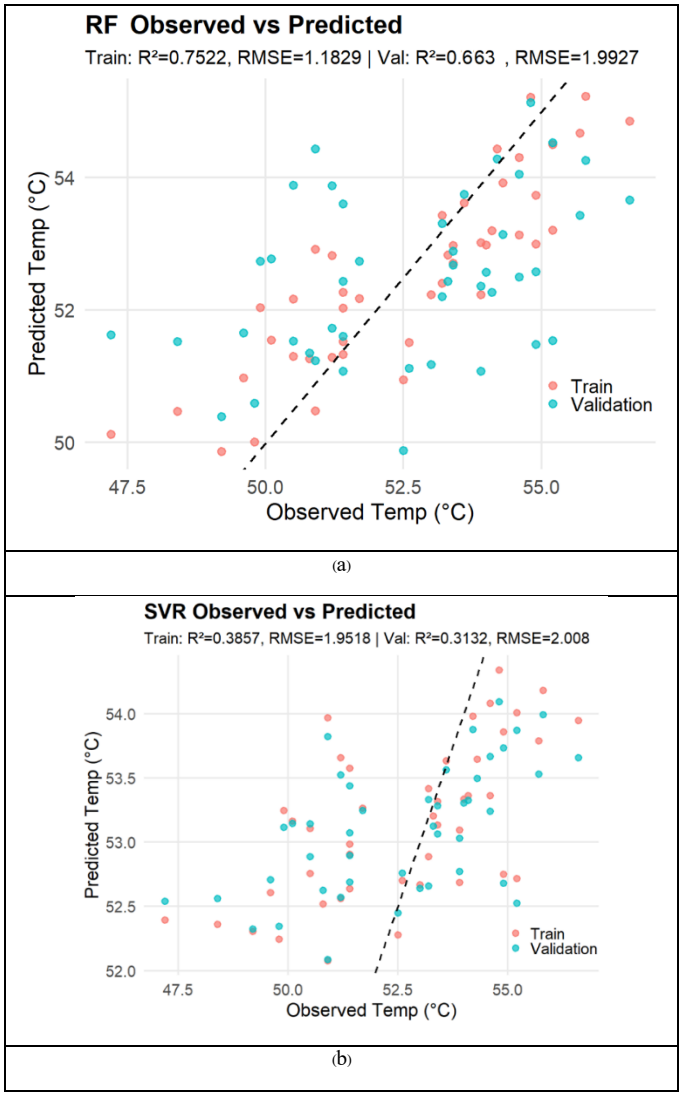


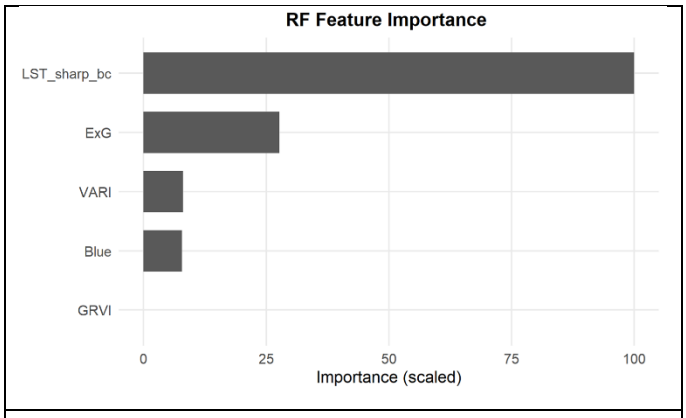
Figure 2: Scatter plots of actual versus predicted surface temperatures for (a) RF, and (b) SVR, models, showing training (green) and testing (red) datasets with the 1:1 reference line

2. Feature Importance Analysis

The permutation-based feature importance analysis for the RF model is presented in Figure 3: Among all predictors, the sharpened and bias-corrected Landsat-derived LST (LST_sharp_bc) was identified as the dominant variable, contributing substantially more than the UAV-derived spectral indices.

Among the RGB-based predictors, Excess Green (ExG) exhibited the highest relative importance, followed by VARI, Blue reflectance, and GRVI. The prominence of LST_sharp_bc highlights the critical role of thermally constrained satellite information in preserving physically meaningful temperature gradients. Meanwhile, the contribution of ExG and VARI suggests that subtle reflectance variations associated with surface material properties and minor vegetation presence can refine sub-pixel thermal heterogeneity.

The comparatively lower importance of individual RGB bands indicates that raw reflectance values alone are insufficient to explain temperature variability without the thermally informed predictor. This finding supports the effectiveness of the fusion-based modeling framework, where satellite-derived thermal consistency is enhanced rather than replaced by high-resolution UAV spectral detail.



(a)

Figure 3: Top five most important predictors identified by permutation-based feature importance for (a) Random Forest (RF) models

3. Feature Importance Analysis

The overall UAV–Landsat thermal downscaling workflow is illustrated in Figure 4. As shown in Figure 4:(a), the fine-scale UAV imagery domain provides high-resolution (0.2 m) RGB reflectance information, enabling detailed characterization of surface material variability and micro-scale heterogeneity along the road corridor. In contrast, Figure 4:(b) represents the coarse-resolution satellite thermal domain, where Landsat-derived Land Surface Temperature (LST) preserves physically consistent thermal support at 100 m spatial resolution.

The integration of these two domains is conceptualized in Figure 4:(c), where UAV-derived spectral predictors are combined with bias-corrected Landsat LST within a machine learning regression framework to generate high-resolution Road Surface Temperature (RST) maps. This fusion-based approach ensures that fine spatial details from UAV imagery are constrained by the thermally meaningful satellite signal, thereby preserving large-scale temperature consistency while enhancing sub-pixel spatial variability. The framework explicitly addresses the spatial–temporal resolution gap between satellite and UAV platforms, enabling physically coherent and spatially refined thermal mapping of urban road infrastructure.

(a)

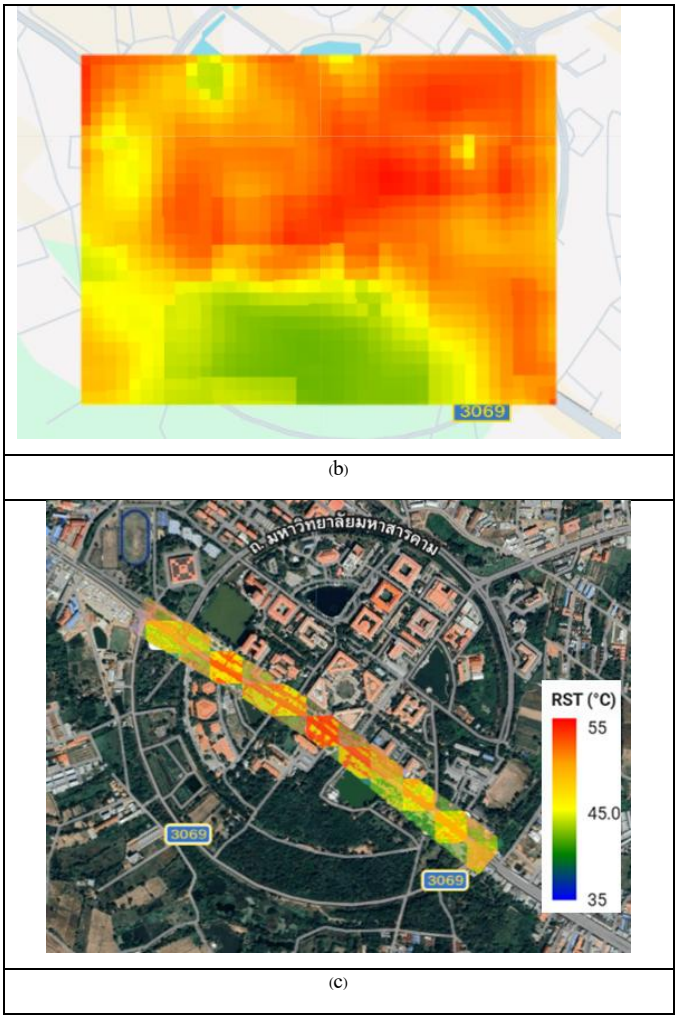


Figure 4: UAV–Landsat thermal downscaling framework: (a) fine-scale UAV imagery domain, (b) coarse satellite thermal domain, and (c) Road Surface Temperature Mapping Using UAV Data and Landsat Thermal Observations

Discussion

The results demonstrate that the Random Forest (RF) model substantially outperformed the Support Vector Regression (SVR) model in estimating road surface temperature (RST), as evidenced by higher R^2 and lower RMSE values (Table 1; Figure 2:). The RF model achieved stable generalization performance under five-fold cross-validation, with a limited difference between training and testing R^2 ($\Delta R^2 = 0.09$), indicating controlled overfitting and robust predictive capability. This finding is consistent with previous studies reporting that ensemble-based machine learning approaches are particularly effective for modeling nonlinear thermal behavior across heterogeneous surfaces (Dai et al., 2025; Wu et al., 2022; Su et al., 2025). Urban road environments are characterized by complex radiometric responses influenced by asphalt composition, surface aging, traffic-induced heat emissions, and surrounding land cover interactions (Kawakubo et al., 2023; Henn et al., 2024), making nonlinear ensemble methods more suitable than kernel-based regressions under limited sample conditions.

The scatter plots (Figure 2:) further confirm the superiority of RF, showing strong alignment between observed and predicted temperatures along the 1:1 reference line, particularly within the dominant temperature range. In contrast, SVR exhibited weaker sensitivity to higher temperature values

and greater dispersion across the full range, suggesting reduced capability in capturing localized thermal variability. Similar patterns have been observed in multi-sensor fusion studies where tree-based ensemble models outperform single-kernel regressors when applied to complex urban microclimate modeling (Su et al., 2025). While earlier thermal downscaling investigations largely focused on vegetated or agricultural systems (Wu et al., 2022; Aboutalebi et al., 2022), the present study extends this methodological framework to impervious urban road infrastructure, which presents distinct thermal and spectral characteristics due to its material properties and anthropogenic influences (Kawakubo et al., 2023; Zargari et al., 2024).

The permutation-based feature importance analysis (Figure 3) highlights that the sharpened and bias-corrected Landsat-derived LST (LST_sharp bc) is the dominant predictor in the RF model. This result emphasizes the importance of preserving physically meaningful thermal support during downscaling. Satellite-derived LST products from Landsat provide temporally consistent and physically grounded thermal measurements but are spatially constrained by coarse resolution (Bahi et al., 2025; Aboutalebi et al., 2022). The present findings demonstrate that UAV-derived spectral indices primarily refine spatial heterogeneity within thermal pixels rather than replace the coarse-scale thermal signal. Among UAV predictors, Excess Green (ExG) and VARI contributed secondary explanatory power, likely reflecting subtle variations in surface reflectance associated with material differences and minor vegetation presence along road edges. These observations align with previous research indicating that surface reflectance variability and material composition significantly influence localized temperature distribution (Andriambololonaharisoamalala et al., 2025; Henn et al., 2024). However, the relatively lower importance of raw RGB bands confirms that reflectance information alone is insufficient for accurate thermal estimation without a thermally constrained satellite predictor.

Despite the strong performance of the proposed framework, several methodological considerations must be acknowledged. Landsat TIRS Band 10 has a native spatial resolution of 100 m, which defines the physical support of the thermal measurement. Downscaling cannot generate new independent thermal information beyond this scale. Instead, the approach models sub-pixel variability through statistical relationships between aggregated UAV spectral indices and Landsat LST. Consequently, the 0.20 m temperature outputs represent spatially refined estimations constrained by coarse-resolution thermal observations rather than direct high-resolution thermal measurements. The underlying assumption that spatial variations in reflectance are physically linked to temperature variability—is supported by surface energy balance theory and prior remote sensing studies (Dai et al., 2025; Wu et al., 2022; Su et al., 2025), but may not fully capture complex radiative interactions in mixed urban materials. The bias correction procedure ensures conservation of thermal energy by forcing the mean of downscaled predictions within each 100 m pixel to equal the original Landsat LST, thereby preserving large-scale consistency. However, this does not eliminate uncertainty associated with intra-pixel temperature distribution. Furthermore, the reported R^2 and RMSE values reflect predictive accuracy at the native 100 m thermal scale and should not be interpreted as evidence of true 0.20 m thermal measurement accuracy.

Nevertheless, the methodological design of this study remains both physically and statistically defensible. Model training was conducted strictly at the native thermal resolution. UAV predictors were aggregated to match the thermal support of Landsat observations prior to regression; no artificial thermal resampling was performed. In addition, block-wise bias correction preserved energy consistency, while model validation using five-fold cross-validation ensured predictive robustness. Compared with direct pixel resampling approaches that may distort the physical meaning of thermal data, the fusion-driven framework presented here provides a coherent solution to the spatial-temporal resolution gap identified in UHI monitoring research (Bahi et al., 2025; Aboutalebi et al., 2022). By integrating temporally consistent satellite thermal observations with high-resolution UAV reflectance

information, the approach improves the spatial interpretability of urban pavement thermal patterns while preserving the integrity of the original satellite measurements.

Beyond methodological contributions, the findings of this study also provide important implications for urban heat mitigation and infrastructure management. Road networks constitute a major proportion of impervious urban surfaces and significantly contribute to localized heat accumulation. The ability to generate spatially refined road surface temperature maps can assist urban planners and municipal authorities in identifying thermal hotspots along transportation corridors. Such information may support evidence-based decisions regarding pavement material selection, reflective pavement implementation, and roadside vegetation planning aimed at reducing heat accumulation. Furthermore, detailed RST mapping can help policy makers prioritize targeted cooling interventions in areas with dense road infrastructure and high pedestrian exposure.

Overall, this study contributes to UHI-oriented thermal monitoring by demonstrating that multi-sensor fusion techniques—previously emphasized in vegetation and agricultural applications (Wu et al., 2022; Aboutalebi et al., 2022)—can be effectively adapted to urban road infrastructure. By providing spatially refined yet physically consistent temperature estimates, the proposed framework offers a valuable analytical tool for supporting urban heat mitigation strategies and climate-resilient infrastructure planning in rapidly urbanizing environments.

Conclusion

This study developed and evaluated a fusion-driven thermal downscaling framework that integrates high-resolution UAV RGB imagery with Landsat-derived Land Surface Temperature (LST) to generate spatially refined Road Surface Temperature (RST) maps in an urban environment. By combining thermal sharpening with machine learning regression, the framework effectively bridges the spatial–temporal resolution gap between satellite and UAV platforms.

The Random Forest (RF) model demonstrated superior predictive performance compared to Support Vector Regression (SVR), achieving stable generalization under five-fold cross-validation (Test $R^2 = 0.66$; RMSE = 1.99°C). The limited difference between training and testing performance confirms the robustness of the proposed modeling approach for heterogeneous urban road surfaces. Feature importance analysis further revealed that the sharpened and bias-corrected Landsat-derived LST was the dominant predictor, while UAV-derived spectral indices such as Excess Green (ExG) and VARI contributed to refining sub-pixel thermal variability. These findings confirm that satellite thermal support remains the primary driver of temperature estimation, whereas UAV spectral information enhances spatial detail without compromising physical consistency.

Importantly, the proposed framework does not create independent fine-scale thermal measurements beyond the native 100 m resolution of Landsat TIRS. Instead, it provides a statistically defensible and physically constrained refinement of thermal patterns, preserving energy consistency through block-wise bias correction. Compared with direct resampling approaches, this method offers a more reliable strategy for improving spatial interpretability of urban thermal environments.

By extending multi-sensor fusion techniques—commonly applied in vegetation and agricultural studies— to impervious urban road infrastructure, this research advances UHI-oriented thermal monitoring and pavement thermographic analysis. The ability to generate high-resolution yet physically consistent RST maps supports data-driven decision-making for pavement material evaluation, urban heat mitigation strategies, and climate-resilient infrastructure planning under intensifying Urban Heat Island

conditions.

Future work should explore physically based thermal sharpening methods and incorporate additional multispectral or hyperspectral UAV data to further improve sub- pixel temperature characterization and strengthen physical interpretability.

Acknowledgment

This research project was financially supported by Faculty of Engineering, Mahasarakham University.

References

Aboutalebi, M., Torres-Rua, A., McKee, M., & Kustas, W. P. (2022). Downscaling UAV land surface temperature using a coupled wavelet–machine learning–optimization algorithm and its impact on evapotranspiration. *Irrigation Science, 40*, 1–18.

Andriambololonaharisoamalala, R. R., Helmholz, P., Bulatov, D., et al. (2025). Downscaling of urban land surface temperatures using geospatial machine learning with Landsat 8/9 and Sentinel-2 imagery. *Remote Sensing, 17*(14), 553-574.

Bahi, H., Bounoua, L., Sabri, A., Bannari, A., Malah, A., & Rhinane, H. (2025). A new thermal fusion method to downscale land surface temperature to finer spatial resolution using Sentinel-MSI and Landsat-OLI/TIRS imagery. *Remote Sensing Applications: Society and Environment, 37*, 101519. https://doi.org/10.1016/j.rsase.2025.101519

Dai, R., Takenaka, K., Xu, X., Shen, Y., & Liu, Y. (2024). UAV-based fine-scale daily relationship characterization of 3D urban surface thermal dynamics. *Science of the Total Environment, 923*, 171466. https://doi.org/10.1016/j.scitotenv.2024.171466

Henn, K. A., George, S. A., & Carlson, J. E. (2024). Surface heat monitoring with high-resolution UAV thermal imagery in urban environments. *Remote Sensing, 16*(5), Article number not specified.

Kawakubo, S., Toratani, T., & Sato, M. (2023). Visualization of urban roadway surface temperature by infrared imagery and deep learning. *Remote Sensing Applications: Society and Environment, 31*, Article 100972 (if available).

Su, Q., Guo, X., Yang, Y., et al. (2025). Quantification of MODIS land surface temperature downscaling via machine learning algorithms. *Remote Sensing, 17*(14), Article number not specified.

Sun, Y., Li, T., Gong, W., et al. (2019). Quantifying the effects of urban form on land surface temperature using random forest and OLS. *Remote Sensing, 11*(8), Article 959 (if available).

Wang, F., Yang, S., Yang, W., Yang, X., & Jianli, D. (2019). Comparison of machine learning algorithms for soil salinity predictions in three dryland oases located in Xinjiang Uyghur Autonomous Region of China. *European Journal of Remote Sensing, 52*, 256–276.

Wei, Y., Shi, Z., Biswas, A., Yang, S., Ding, J., & Wang, F. (2020). Updated information on soil salinity in a typical oasis agro-ecosystem and desert-oasis ecotone: Case study

along the Tarim River, China. *Science of the Total Environment, 716*, Article 137030 (if available).

Wu, J., Wang, Y., Li, W., Liu, Q., & Cui, H. (2022). Downscaling land surface temperature: A framework based on multi-scale feature embedding and residual learning. *Remote Sensing of Environment, 281*, Article 113240 (if available).

Zargari, M., Haidinia, R., & Montazeri, H. (2024). Climatic comparison of surface urban heat island using day/night thermal data: Case of Tehran. *Scientific Reports, 14*, Article 643. https://doi.org/10.1038/s41598-023-50757-2

COMPARATIVE BIOMECHANICAL EFFECTIVENESS OF OPEN-WEDGE AND CLOSED-WEDGE HIGH TIBIAL AND DISTAL FEMORAL OSTEOTOMIES IN KNEE OSTEOARTHRITIS

Lin Lin Lett[1] and Nattapon Chantarapanich[2*]

[1,2]Department of Mechanical Engineering, Faculty of Engineering at Sriracha, Kasetsart University Sriracha Campus, Chonburi, Thailand

[*]**Corresponding Author, E-mail**: nattapon@eng.src.ku.ac.th

Abstract

This study evaluated the biomechanical effects of High Tibial Osteotomy (HTO) and Distal Femoral Osteotomy (DFO) on tibiofemoral load distribution in both varus and valgus knee deformities using finite element analysis (FEA). Twelve three-dimensional (3D) whole-limb FE models were reconstructed from computed tomography (CT) data under identical geometry, material properties, and loading conditions to represent preoperative hip–knee–ankle (HKA) alignments of varus (−6°HKA) and valgus (+5°HKA), and their corresponding postoperative correction states. Varus and valgus deformities were corrected to neutral alignment (0° HKA), with varus cases additionally evaluated under (+5° HKA) valgus overcorrection. All models were subjected to a 348 N load applied at the femoral head, representing approximately half of body weight during double-leg standing. Prior to correction, the varus knee demonstrated dominant medial compartment loading (72.7%:27.3%). MOWHTO shifted the load toward the lateral compartment (34.81%:65.19% at 0° HKA; 31.11%:68.89% at +5° HKA), whereas LCWHTO produced a smaller shift (36.2%:63.8%). DFO produced more moderate redistribution, with MOWDFO yielding ratios of 38.13%:61.87% at 0° HKA and 30.76%:69.24% at +5° HKA, and LCWDFO showing 37.93%:62.07%. In valgus alignment, lateral compartment loading predominated (31.9%:68.1%). After correction to 0° HKA, LOWDFO and MCWDFO showed similar distributions (31.85%:68.15% and 30.99%:69.01%). In contrast, tibial-based corrections further increased lateral dominance, with LOWHTO yielding 25.64%:74.36% and MCWHTO yielding 26.97%:73.03%. Overall, varus correction was most effective with HTO, whereas valgus correction was better addressed with DFO.

Keywords: Knee Osteoarthritis, High Tibial Osteotomy, Distal Femoral Osteotomy, Finite Element Analysis, HKA Alignment

Introduction

Osteotomy procedures, first introduced in the 16th century (Dabis et al., 2017) and significantly refined from the 19th to the 21st centuries, have become standard joint-preserving treatments for unicompartmental knee osteoarthritis (KOA) (Ferrera & Menetrey, 2022). KOA is a progressive joint disease and one of the main causes of disability worldwide. It is characterized by cartilage breakdown, bone remodeling, inflammation, and chronic pain. The condition is becoming more common due to aging, obesity, sedentary lifestyles, and genetic or metabolic factors, which reduce mobility, limit daily activities, and lower the quality of life. Treatment usually begins with conservative approaches such as weight management, physiotherapy, pharmacologic therapy, including nonsteroidal anti-inflammatory drugs (NSAIDs), and intra-articular injections to relieve symptoms and slow disease progression. However, in advanced stages, when pain and functional limitations persist, surgical intervention may be required. The two main surgical options are HTO and total knee arthroplasty (TKA), with TKA widely regarded as an effective treatment for end-stage KOA when conservative measures fail (Silva Araújo et al., 2025). TKA provides significant pain relief and functional improvement; however, issues such as tibial component malalignment, rotational errors, and abnormal biomechanical loading remain major causes of implant failure, joint instability, and revision surgery (Diconi et al., 2025). Consequently, joint-redistribution and such as HTO and DFO remain important alternatives, particularly for younger and more active patients, as they preserve the natural joint structure, maintain more physiological load distribution, and delay the need for joint replacement.

KOA has been a major focus of orthopedic research since the mid-20th century owing to its high global prevalence and substantial impact on pain, function, and disability (Kellgren & Lawrence, 1957). As a degenerative condition, it primarily affects older adults, with progressive cartilage loss most commonly occurring in the medial tibiofemoral compartment, leading to joint space narrowing and varus malalignment, while lateral compartment degeneration results in valgus deformity. From a biomechanical perspective, KOA represents a progressive mechanical failure in which cartilage deterioration disrupts lower-limb alignment, increases joint stress, and accelerates structural degeneration (Chantarapanich et al., 2009; Kim et al., 2024).Globally, KOA represents approximately 80% of all osteoarthritis cases and affects about 16% of adults, with prevalence rising sharply after age 40 and occurring more frequently in women (Cui et al., 2020). Nearly 45% of women over 65 report KOA

symptoms, and about 70% show radiographic evidence of disease. Women often report more severe pain, greater functional limitations, emotional distress, and reduced quality of life. KOA's chronic and progressive nature may also contribute to depression, sleep disturbances, and reduced vitality (Asgaonkar & Ghugare, 2020). Although joint replacement provides clinical improvement, its use in younger, active individuals remains debated due to implant wear and declining long-term survivorship (Bergstein et al., 2024).

The knee is a mechanically complex joint that functions as a modified hinge, allowing flexion, extension, and limited rotation, while articular cartilage facilitates smooth motion and load distribution (Yan et al., 2024). Lower limb alignment is commonly assessed using the HKA angle, a crucial parameter for diagnosing deformities and planning surgical correction. Manual HKA measurement, however, remains prone to variability (Kim et al., 2024). Alignment is typically classified as neutral (HKA = 180°), varus (HKA < 180°), or valgus (HKA > 180°) (Wu et al., 2022, Xu et al., 2024a). Varus deformity involves medial deviation of the mechanical axis, increasing medial tibiofemoral loading and cartilage stress (Kim et al., 2024). This aligns with earlier findings by Chantarapanich et al. (2009), who reported increased medial compartment stress in varus knees. HTO is a joint-preserving procedure that corrects varus malalignment by shifting the mechanical axis laterally, thereby reducing medial compartment loading. It is typically indicated for young, active patients with medial compartment KOA. However, surgical outcomes depend on correction accuracy, as both under- and overcorrection adversely affect long-term results (Zhang et al., 2023; Trad et al., 2018).

Medial open-wedge high tibial osteotomy (MOWHTO) is the most widely used technique for varus malalignment. It creates a medial opening gap in the proximal tibia to laterally shift the mechanical axis (MA), thereby reducing medial compartment contact force and cartilage stress. This technique enables precise alignment correction, preserves bone stock, and minimizes soft-tissue disruption, resulting in favorable biomechanical outcomes (Liu et al., 2019). MOWHTO provides superior medial compartment unloading and alignment accuracy compared to lateral closed-wedge HTO (LCWHTO), although LCWHTO offers greater structural stability. However, MOWHTO carries a risk of lateral hinge fracture (LHF), which may compromise construct stability and delay healing (Kang et al., 2020). Furthermore, overcorrection in OWHTO is associated with increased complication rates without improved functional outcomes, and comparable results between neutral and valgus overcorrection indicate that the optimal alignment target remains uncertain (Xu et al., 2024b). In contrast, LCWHTO corrects varus alignment by removing a lateral bone wedge, producing a compressive osteotomy that enhances inherent stability. However, it requires more extensive soft-tissue dissection, increases the risk

of cortical fracture, and is less effective in reducing medial compartment loading and knee adduction moment (Sun et al., 2017; Lee et al., 2018; Liu et al., 2019).Therefore, appropriate technique selection, including OWHTO or tibial condylar valgus osteotomy (TCVO), is critical for maintaining joint stability and optimizing load distribution (Higuchi et al., 2024).

Valgus deformity, characterized by a mechanical axis passing lateral to the knee center, increases lateral tibiofemoral loading and accelerates lateral cartilage degeneration (Ismailidis et al., 2023). Although less prevalent than varus deformity, valgus malalignment poses greater biomechanical and surgical challenges due to associated bony and soft-tissue abnormalities (Wu et al., 2022). DFO is the preferred joint-preserving procedure for young, active patients with valgus alignment and lateral compartment KOA, as most valgus deformities originate from the distal femur (Ferrera & Menetrey, 2022; Ismailidis et al., 2023). DFO can be performed as a medial closing wedge (MCWDFO) or lateral opening wedge (LOWDFO). MCWDFO and LOWDFO both restore alignment effectively but exhibit different biomechanical characteristics. MCWDFO provides greater construct stability and allows earlier weight bearing, whereas LOWDFO enables controlled angular correction but may require bone grafting (Wylie & Maak, 2016; Ismailidis et al., 2023).FE studies show that slight varus overcorrection (0°–10°) in LOWDFO reduces lateral compartment stress but increases medial compartment loading, indicating load redistribution (Wu et al., 2022). In addition, MCWDFO and LOWDFO differ in hinge mechanics and fracture risk (Meisterhans et al., 2023), al., 2023; Ferrera & Menetrey, 2022). However, the optimal correction angle remains unclear, with neutral alignment associated with more balanced tibiofemoral load distribution (Bolcos et al., 2018; Li et al., 2020; Wu et al., 2022).

Despite extensive research on individual techniques, including outcomes using different correction angles, patient-specific models, and varying experimental or computational conditions, these comparisons are often conducted separately for tibial or femoral osteotomies and for open- or closed-wedge techniques, rather than under identical conditions. As a result, biomechanical differences remain unclear, and direct comparisons of open- and closed-wedge osteotomies at both tibial and femoral levels under identical varus and valgus conditions still reamin. Therefore, the present study aims to compare open- and closed-wedge HTO and DFO using identical 3D CT-based FE models to provide objective biomechanical evidence for optimal technique selection. Accordingly, the study objectives are as follows:

Research Objectives

This study has two objectives:

1. To quantify changes in medial and lateral tibiofemoral load distribution in varus and valgus knee alignments before and after correction, including correction to neutral alignment (0° HKA) and additional valgus overcorrection (+5° HKA) in selected varus cases, using FEA simulation under static double-leg standing loading conditions.
2. To compare the biomechanical effectiveness of HTO and DFO in redistributing tibiofemoral loads in varus and valgus knee deformities, under identical FE model conditions, to determine whether tibial- or femoral-based correction is more appropriate.

Literature Review

The biomechanics of knee alignment has been extensively investigated using *in silico*, *in vitro*, and *in vivo* methodologies, each offering unique insights into tibiofemoral load transmission and osteoarthritis progression while also presenting distinct methodological challenges. Radiographic severity of KOA is commonly assessed using the Kellgren–Lawrence (KL) classification system, first introduced in 1957 and later endorsed by the World Health Organization (WHO) in 1961 (Asgaonkar & Ghugare, 2020). From a methodological perspective, prior studies can be grouped into computational (*in silico*), clinical (*in vivo*), and experimental (*in vitro*) approaches.

In silico approaches have played a foundational role in understanding mechanical alterations *In silico* approaches have played a foundational role in understanding mechanical alterations associated with malalignment. Chantarapanich et al. (2009) reconstructed CT- and radiograph-based whole-limb numerical models and demonstrated that varus alignment produces substantially increased medial tibiofemoral stress compared with neutral alignment. Trad et al. (2018) further emphasized that computational predictions of cartilage stress, meniscal load sharing, and osteotomy outcomes are highly sensitive to lower-limb alignment, subject-specific anatomical variations, and soft-tissue modeling assumptions. A major challenge of *in silico* modeling, therefore, lies in accurately replicating complex anatomical geometries, material properties, and physiological boundary conditions, as these assumptions can directly influence predicted joint reaction forces. More recent patient-specific FEA illustrate the influence of alignment correction on joint biomechanics. Wu et al. (2022) reported that correction toward neutral alignment optimizes medial–lateral stress distribution, whereas excessive valgus overcorrection elevates medial compartment stress. Kang et al. (2020) and Pan et al. (2023) demonstrated that surgical factors such as lateral hinge fracture configuration and bone graft position in (MOWHTO) significantly affect construct stability, ligament loading, and interfragmentary

micromotion. Although FEA is beneficial for estimating internal stresses and forces that are inaccessible through experimental means, model validation remains challenging because of the difficulty of directly measuring in vivo tibiofemoral forces.

In vivo studies, by contrast, focus on physiological joint behavior, including neuromuscular control, gait mechanics, and long-term biological responses, yet direct measurement of knee joint forces is rarely feasible (Ferrera & Menetrey, 2022; Ismailidis et al., 2023; Zhang et al., 2023). Ethical concerns, technical constraints, and tissue disruption associated with implantable sensors have limited clinical research primarily to indirect assessment of joint loading (Chantarapanich et al., 2009; Trad et al., 2018; Wu et al., 2022). Follow-up studies indicate that TKA is effective in treating valgus deformity; however, valgus knees are associated with higher complication and revision rates compared with varus knees. Common complications include ligament instability, patellar subluxation or fracture, and loosening prosthesis, and longer operative time and hospital stay may further increase revision risk (Xu et al., 2024a). Given these challenges, comprehensive evaluation and long-term follow-up are essential to guide clinical decision-making and postoperative rehabilitation. In this context, balance control has been investigated in patients with varus deformity before and after HTO. It has been suggested that correction of lower limb alignment through HTO can improve joint function and enhance postoperative postural stability, potentially restoring impaired preoperative balance control to near-normal levels (Zhang et al., 2023). (MCWDFO) is often preferred because it provides reliable bone healing, strong stability, and fewer hardware complications. Although the lateral opening-wedge technique allows easier adjustment during surgery, it has been associated with delayed healing and higher hardware irritation. DFO has a reported reoperation rate of 25–40%, mainly due to hardware removal or later conversion to total knee arthroplasty. However, in properly selected younger patients, outcomes are generally favorable. Overall, DFO effectively corrects valgus alignment, reduces pain, and improves knee function (Wylie & Maak, 2016).

Elyasi et al. (2021) found that (OWHTO) can increase posterior tibial slope, decrease patellar height, and cause medial collateral ligament tightness due to wedge opening and soft-tissue changes. Proper preoperative planning and biomechanical analysis are important to reduce these complications. HTO is an important treatment for medial compartment osteoarthritis with varus deformity, as it shifts the load-bearing axis laterally to improve alignment. Successful outcomes depend on understanding its biomechanical effects on gait, joint mechanics, and contact forces. However, current research remains limited, particularly in gait analysis, joint kinematics, and contact mechanics. Combining

musculoskeletal modeling and computational analysis may improve understanding of patient-specific biomechanics, optimize correction accuracy, and reduce complications (Liu et al., 2019). Osteotomies around the knee are reliable surgical options for young, active patients with unicompartmental osteoarthritis or ligament instability. Successful outcomes depend on proper patient selection, accurate preoperative planning, and precise surgical technique. Advances in software, 3D technology, and navigation systems are expected to further improve surgical planning and intraoperative correction accuracy (Ferrera & Menetrey, 2022).

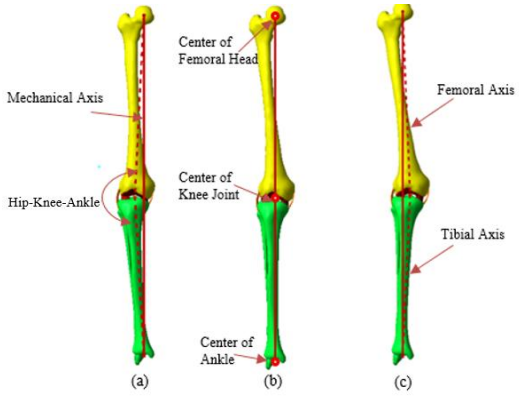
Among the two principal techniques of HTO, (MOWHTO) is the most frequently performed due to its high correction accuracy and effective medial unloading, particularly when combined with angular-stable fixation systems that reduce micromotion and support predictable bone healing (Liu et al., 2019; Carranza et al., 2025; Chen et al., 2024; Kang et al., 2020). In (MOWHTO), normal correction targets neutral alignment to restore the mechanical axis, improve function, and reduce symptoms. Neutral alignment has been shown to enhance knee function scores and is effective across different age groups. Overcorrection shifts alignment beyond neutral to optimize stress distribution and prevent recurrence, but it requires careful planning to avoid complications and ensure long-term success (Osmani et al., 2024). In contrast, (LCWHTO) may increase soft-tissue disruption and risk of cortical fracture and offers less favorable control of varus moments and lateral instability, despite inherent structural stability (Liu et al., 2019). Even with advancements in 3D alignment planning, challenges such as hinge instability, alterations in posterior tibial slope, and delayed union persist, emphasizing the need for precise surgical planning (Nakamura et al., 2024; Liu et al., 2019). HTO is a widely utilized joint-preserving intervention for young, active patients with medial compartment KOA, aiming to shift the mechanical axis laterally, reduce medial loading, enhance function, and delay the need for total knee arthroplasty (Elyasi et al., 2021; Zhang et al.,2023).

Complementing computational research, *in vitro* biomechanical investigations provide controlled experimental validation. Using synthetic femur models subjected to servo-hydraulic mechanical testing, Karkosch et al. (2024) demonstrated that (MOWDFO) stabilized with Tomofix plates achieves axial and torsional stability comparable to the (LOWDFO) technique. Cadaveric analyses by Li et al. (2020) using ultra-low-pressure Fuji Prescale films similarly revealed progressive increases in medial tibial plateau stress with worsening varus deformity and corresponding increases in lateral stress in valgus deformity following malunited tibial fractures. Although *in vitro* methods provide highly repeatable force measurements, they are limited by simplified loading conditions, lack of dynamic muscle forces, and absence of long-term physiological adaptation. Nonetheless, these results confirm

that coronal malalignment significantly alters tibiofemoral load distribution.

For valgus deformity, two major of DFO techniques are commonly applied: (LOWDFO) and (MCWDFO) (Wylie & Maak, 2016; Ferrera & Menetrey, 2022; Ismailidis et al., 2023). LOWDFO allows controlled coronal-plane correction and may be beneficial when significant angular adjustment or leg-length restoration is required (Wylie & Maak, 2016; Ferrera & Menetrey, 2022). However, its success depends on maintaining medial hinge integrity; larger corrections markedly increase hinge strain, implant stress, and fracture risk (Wu et al., 2022; Meisterhans et al., 2023). MCWDFO, involving wedge resection and direct bone contact, provides greater initial stability, reduced hinge complications, and faster weight-bearing progression (Wylie & Maak, 2016; Ismailidis et al., 2023; Duethman et al., 2019). Clinical evidence suggests MCWDFO is generally favored due to more reliable union, lower complication rates, and enhanced stability (Ferrera & Menetrey, 2022; Ismailidis et al., 2023). Nonetheless, selection of the appropriate technique must be individualized, considering deformity origin, correction magnitude, limb-length considerations, and surgeon expertise (Wylie & Maak, 2016; Ferrera & Menetrey, 2022).

Across all investigative approaches, direct *in vivo* measurement of tibiofemoral forces remains the most challenging, whereas *in vitro* methods enable controlled testing but lack full physiological realism, and *in silico* modeling can evaluate internal stresses and joint force transmission but depends on modeling assumptions. However, existing studies are often conducted separately, providing insufficient direct comparison between tibial and femoral corrections as well as between open- and closed-wedge techniques under identical conditions. Previous studies using *in vivo*, *in vitro*, and *in silico* approaches consistently demonstrate a biomechanical tendency in which varus alignment increases medial loading, while valgus alignment shifts load laterally. However, variations in methodology, differences in modeling assumptions, and challenges in mechanical model validation, in addition to the absence of direct comparison between osteotomy techniques under consistent conditions, indicate that the optimal correction strategy remains insufficient.



Picture 1: Coronal-plane lower limb alignments illustrating (a) varus malalignment, (b) neutral mechanical axis, and (c) valgus malalignment. The red line represents (HKA) angle.

Research Methodology

This protocol of the study was approved for exemption by the Kasetsart University Research Ethics Committee. This study adopted a 3D FE modeling framework to investigate the biomechanical effects of osteotomy technique, correction level, and alignment strategy on tibiofemoral load distribution in knees with coronal-plane malalignment. 3D computational knee and lower-limb geometries were reconstructed from CT data, after which the mechanical axis was defined, and virtual osteotomy corrections were applied. Both tibial-based and femoral-based osteotomy strategies, incorporating open- and closed-wedge configurations, were evaluated under neutral and valgus overcorrection conditions using a design-of-experiments (DOE) in Table 1. Material properties, boundary conditions, contact definitions, and loading protocols were applied consistently across all simulations to isolate the mechanical effects of alignment correction.

Table 1 Conditions under study

No.	Alignment	Osteotomy Type	Osteotomy Technique	Target HKA Alignment
1	Varus	No osteotomy	N/A	(□6°)
2	Varus	HTO	Medial open-wedge	0°
3	Varus	HTO	Medial open-wedge	(+5°)
4	Varus	HTO	Lateral closed-wedge	0°
5	Varus	DFO	Medial open-wedge	0°
6	Varus	DFO	Medial open-wedge	(+5°)

7	Varus	DFO	Lateral closed-wedge	0°
8	Valgus	No osteotomy	N/A	(+5°)
9	Valgus	DFO	Lateral open-wedge	0°
10	Valgus	DFO	Medial closed-wedge	0°
11	Valgus	HTO	Lateral open-wedge	0°
12	Valgus	HTO	Medial closed-wedge	0°

1. CT Acquisition, Segmentation, and Three-Dimensional Reconstruction

High-resolution CT data of the right lower limb from a 28-year-old female subject (body weight: 71 kg) with a healthy knee joint without injury were acquired using a Philips Spectral CT 7500 system. The scan generated 256 axial Digital Imaging and Communications in Medicine (DICOM) slices extending from the femoral head to the distal tibia and fibula. During image acquisition, the subject was positioned with the limb fully extended and relaxed in a neutral anatomical posture to preserve the physiological HKA alignment, as joint flexion or rotational malalignment can affect mechanical-axis (MA) interpretation and compromise the accuracy of alignment-based FE modelling. CT imaging was performed to obtain accurate bony geometry of the knee joint for subject-specific modeling. Although CT provides high geometric fidelity for osseous structures, it has limited capability in capturing soft tissues due to low radiodensity, and validation of CT-based models remains challenging because joint mechanics is complex (Trad et al., 2018). The DICOM datasets were imported into 3DSlicer (Slicer.org) for image segmentation and 3D reconstruction. Bone segmentation was performed using Hounsfield Unit (HU) thresholding to differentiate tissue densities. Cortical bone was defined within a higher HU range (approximately 643–2012 HU), whereas cancellous bone was identified within a lower HU range (approximately 129–642 HU), allowing separate reconstruction of cortical and cancellous components (Pan et al., 2022). To enhance segmentation accuracy, refinement techniques including region growing, dynamic region growing, Boolean operations, and mask editing were applied to eliminate artefacts and restore thin cortical boundaries, particularly in regions with complex geometry or variable cortical thickness (Pan et al., 2022).Following segmentation, the segmented contours from sequential CT slices were merged to generate to generate 3D surface geometries of the femur, tibia and fibula following established CT-based musculoskeletal modeling workflows (Chantarapanich et al., 2009).

Since soft tissues such as femoral cartilage, tibial cartilage, medial and lateral menisci, and the primary knee ligaments (anterior cruciate ligament, posterior cruciate ligament, medial collateral

ligament, and lateral collateral ligament) are not clearly visible on CT images, they were reconstructed in later modeling stages based on validated anatomical references from previously established FE knee models. Ligaments were included to provide joint stability, whereas muscles and tendons were not explicitly modeled; instead, external loads and moments were applied directly to the model (Trad et al., 2018). Following segmentation, the femur, tibia, and fibula geometries were exported as stereolithography (STL) files and imported into Geomagic Studio for surface refinement. Topological corrections were applied to eliminate non-manifold edges, intersecting triangles, and small surface perforations introduced during segmentation. Surface smoothing and validation were then performed to ensure watertight, high-quality geometries appropriate for FEA, in accordance with established CT-based musculoskeletal modelling workflows. The refined bone models were subsequently transferred to VISI to generate final 3D geometries suitable for MA construction and FE simulations (Chantarapanich et al., 2009) (Pan et al., 2022).

2. Surface Refinement, Mechanical-Axis Construction, and Alignment Characterization

CT-derived surface models of femur, tibia, and fibula were refined in Geomagic Studio to correct segmentation-related artifacts and ensure watertight, anatomically accurate geometries suitable for FEA. Topological repair and targeted manual refinement were applied in complex joint regions to preserve surface curvature and enable reliable contact behavior in subsequent simulations (Kang et al., 2023). The finalized models were then transferred to VISI (Hexagon AB, Sweden) for mechanical-axis evaluation and osteotomy planning. MA determination followed a standardized reference-based method. The hip joint center was identified using least-squares sphere fitting of the femoral head (Mahaisavariya et al., 2002), while the knee and ankle joint centers were defined as the midpoints of the femoral condyles and tibial plafond, respectively.

Picture 2: CT-Based HU Thresholding, Anatomical Segmentation, and Initial 3D Reconstruction of the Lower Limb.

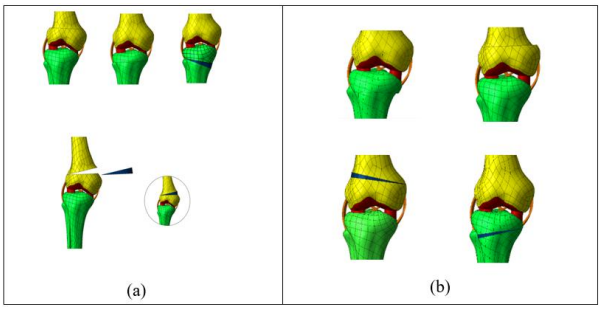
These landmarks were used to construct the (HKA) angle following the procedure described by Chantarapanich (2010), providing an accurate assessment of coronal-plane limb alignment. Based on this approach, the reconstructed limb exhibited a varus alignment of approximately (□6° HKA) and a valgus alignment of approximately (+5° HKA). These quantified alignment conditions established a reliable geometric and biomechanical baseline for subsequent corrective osteotomy simulation and comparative FEA.

3. Osteotomy Modelling and Alignment Correction

Virtual osteotomy planning was performed using VISI software, following established clinical guidelines and correction ranges commonly reported in FE and clinical osteotomy research. For the varus knee (baseline alignment: □6°HKA), HTO was designed with two postoperative alignment targets: restoration to (0° HKA) and overcorrection (+5° HKA). These correction angles are consistent with the classical Fujisawa principle, which recommends positioning the weight-bearing line at 62–70% of the lateral tibial plateau, corresponding to roughly +3° to +6° of valgus, and is widely recognized to effectively unload the medial compartment (Fujisawa et al., 1979). Overcorrected knees are commonly used because patients with a valgus alignment of 3° or greater have been reported to achieve better clinical outcomes and higher postoperative survival rates (Liu et al., 2019). Xu et al. (2024b) showed clinically that patients with a WBLR above 62.5% fall within the valgus overcorrection zone. Correcting varus deformity too much (overcorrection) may not greatly affect clinical results within the first five years (OWHTO). However, it can increase the risk of postoperative complications and may cause greater

compensatory inward movement (adduction) of the hip. Based on this evidence, correcting the (-6°HKA) varus limb to 0° and +5° provided clinically meaningful and biomechanically validated endpoints for the HTO models. For the valgus knee, DFO planning was guided by the findings of Wu et al. (2022), who modelled LOWDFO corrections from 0° to 10° of induced varus (equivalent to 180°–170° HKA). Their results demonstrated that neutral or slight varus (0°–2°) produced balanced medial–lateral stress distributions, whereas stronger varus overcorrection resulted in sharply elevated medial stresses. Therefore, valgus models were corrected only to neutral alignment (0° HKA).

For the varus limb, proximal tibial osteotomies were modeled as MOWHTO and LCWHTO. In (MOWHTO), the osteotomy plane was positioned 35–40 mm below the tibial plateau in the metaphyseal region, extending from the medial cortex toward the lateral metaphysis (Liu et al., 2019; Koh et al., 2019; Pan et al., 2022). Hydroxyapatite, a commonly used material in implant applications, was selected as the simulation material in this study, following evidence that posterior graft support improves medial gap stability and posterior cortical load distribution (Pan et al., 2022). The specifications, such as wedge size and correction angle, have been detailed in previous studies (Liu et al., 2019; Koh et al., 2019; Pan et al., 2022; Wu et al., 2022).



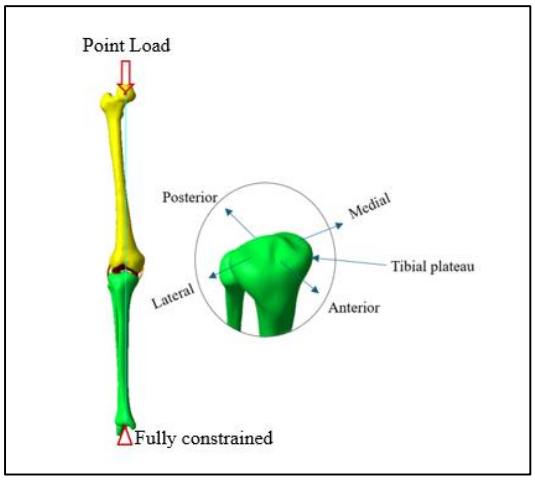
Picture 3: (a) 3D model of varus osteotomy showing neutral and overcorrection alignments; (b)3D modeling of a valgus osteotomy in neutral alignment.

4. Material Properties, Mesh Generation, and Boundary Conditions

All materials in the FE models were assumed to be homogeneous, isotropic, and linearly elastic, following standard assumptions widely used in static knee FEA (Chantarapanich et al., 2009; Trad et al., 2018). Material properties listed in Table 2 for cortical and cancellous bone, cartilage, menisci, ligaments, and hydroxyapatite graft material were assigned based on previously published studies (Chantarapanich et al., 2009; Pan et al., 2022).Meshing was performed with four-node linear tetrahedral

elements were used for meshing, with regional refinement applied to contact interfaces, ligament attachment zones, and osteotomy regions to enhance numerical accuracy without compromising computational efficiency (Donahue et al., 2003; Pena et al., 2006).A maximum element size of □1.5 mm was selected to balance accuracy and computational cost (Donahue et al., 2003; Peña et al., 2006; Chantarapanich et al., 2009; Pan et al., 2022).

Mesh densities ranged from 0.8–1.5 mm for cortical bone, 0.6–1.0 mm for cancellous bone and cartilage, and ~0.6 mm for ligaments. Models were analyzed in Marc Mentat (MSC Software, USA). The distal tibia and fibula were fully constrained, and a vertical load of 348 N (50% body weight) was applied at the femoral head along the mechanical axis to simulate double-leg standing (Chantarapanich, 2010; Wu et al., 2022; Xu et al., 2024b).Frictionless contact was defined between cartilage and menisci, while ligaments were modeled using tied constraints. Nonlinear static analyses were performed to extract medial and lateral meniscal contact forces. Each whole-limb model included articular cartilage and ligaments to preserve realistic alignment-dependent load transfer (Chantarapanich, 2010; Wu et al., 2022). The number of elements ranged from approximately 253,000 to 897,000 depending on the deformity type and osteotomy configuration. All meshes were verified for element quality, ligament attachments, and contact definitions prior to simulation (Trad et al., 2018). All modeling procedures and assumptions were applied consistently across all cases to ensure reproducibility and fair comparison between osteotomy configurations.



Picture 4:3D lower-limb model with applied loading and boundary conditions, highlighting the tibial plateau orientation (anterior, posterior, medial, and lateral)

Table 2 Material Properties

Region	Modulus of Elasticity (MPa)	Poisson's Ratio
Femur, tibia and fibula		
Cortical bone	17,000	0.3
Cancellous bone	600	0.3
Articular cartilage and meniscus	12	0.45
Ligaments		
ACL	345.0	0.4
PCL	345.0	0.4
MCL	332.2	0.4
LCL	345.0	0.4
Hydroxyapatite graft	11000	0.25

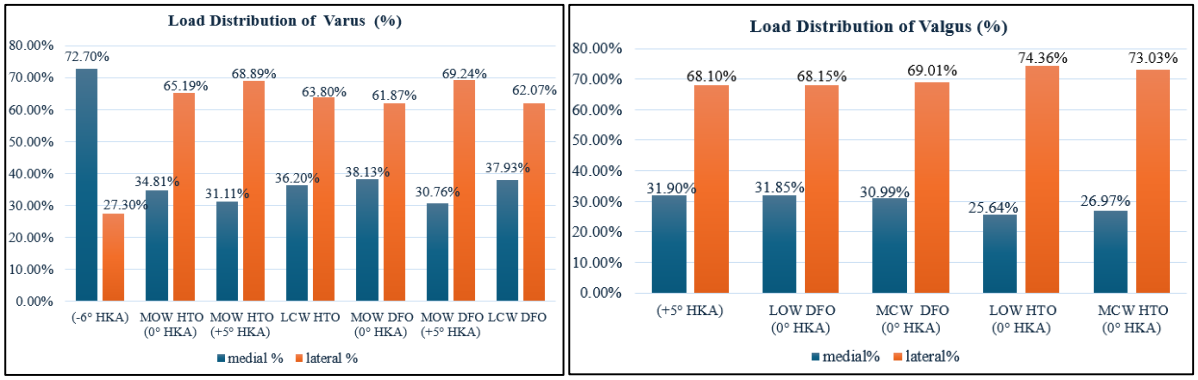
Source: (Chantarapanich et al., 2009; Pan et al., 2022)

Results

The results indicate that varus and valgus malalignments produce opposite medial–lateral tibiofemoral load ratios, and that the effectiveness of correction strongly depends on the anatomical level and technique of the osteotomy. In the pre-correction varus knee (Case 1), loading exhibited a pronounced medial–lateral imbalance toward the medial compartment (medial: 72.7%, lateral: 27.3%). Following correction, MOWHTO (Cases 2–3) achieved the most pronounced redistribution of load. At (0° HKA) correction (Case 2), load sharing shifted to (medial: 34.81%, lateral: 65.19%), while mild valgus overcorrection (Case 3) further altered the distribution to (medial: 31.11%, lateral: 68.89%), indicating unloading of the medial compartment. In comparison, femoral-based correction using DFO (Cases 5–7) produced less pronounced changes, with medial–lateral load sharing ranging from (medial: 38.13%, lateral: 61.87%) at (0° HKA) (Case 5) to (medial: 30.76%, lateral: 69.24%) under valgus overcorrection (Case 6). Although Case 4 reduced medial loading to (medial: 36.2%, lateral: 63.8%), this redistribution was less effective than that achieved with MOWHTO.

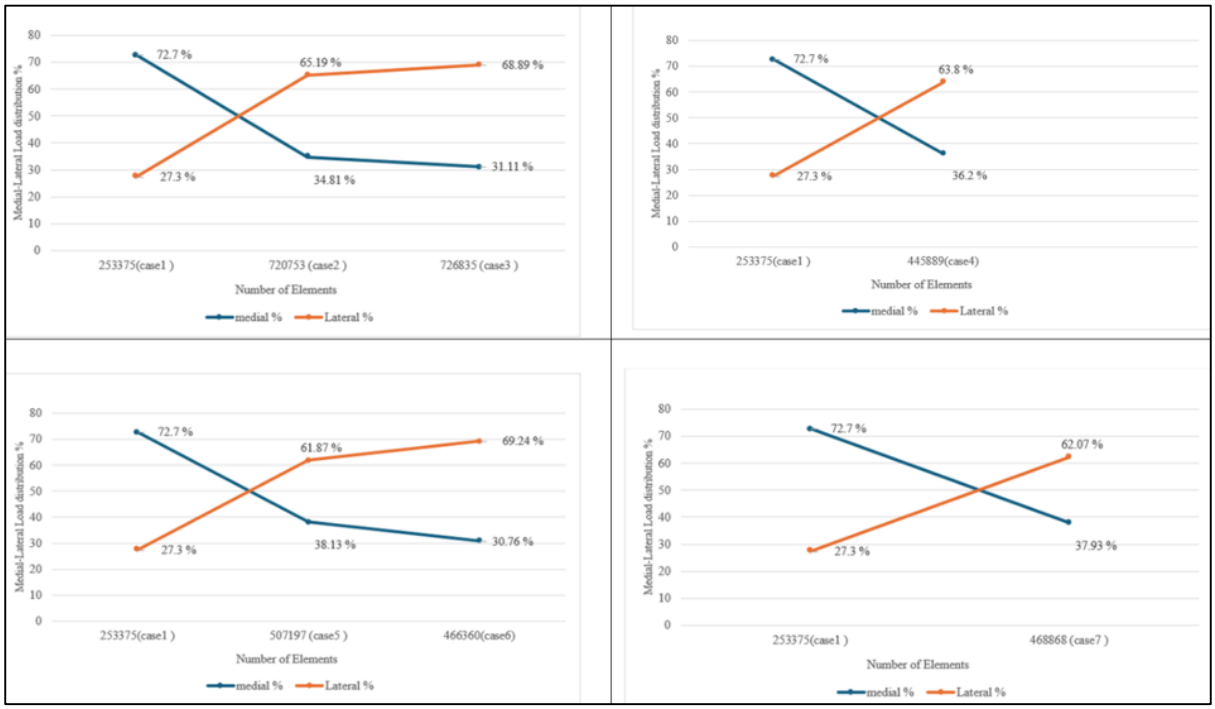
In the pre-correction valgus knee (+5° HKA, Case 8), force transmission was dominantly lateral (medial: 31.9%, lateral: 68.1%). Femoral-based correction produced the most effective redistribution in this

setting: LOWDFO (Case 9) restored a similar load distribution (medial: 31.85%, lateral: 68.15%), while MCWDFO (Case 10) resulted in a comparable pattern (medial: 30.99%, lateral: 69.01%). In contrast, tibial-based corrections in valgus knees (Cases 11–12) further shifted loading laterally to (medial: 25.64%, lateral: 74.36%) and (medial: 26.97%, lateral: 73.03%), respectively. These results indicate that 0° HKA alone is insufficient to substantially redistribute tibiofemoral load in valgus knees, and that slight varus overcorrection may achieve more effective medial–lateral load redistribution. Overall, these findings demonstrate that varus deformity is most effectively corrected using MOWHTO, whereas valgus deformity is more appropriately addressed using femoral-based DFO, with correction at the anatomical origin of the deformity yielding more favorable tibiofemoral load distribution.

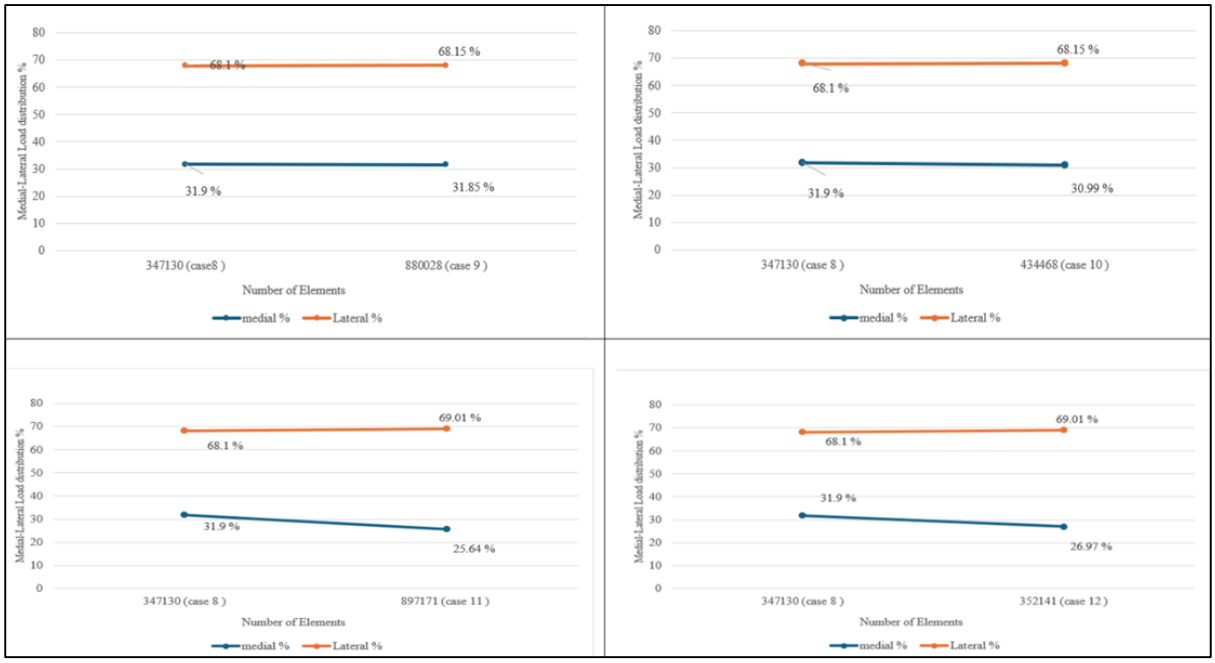


(a) (b)

Picture 5: Results of Tibiofemoral Contact Force Distribution of (a)Varus and (b)Valgus correction



(a)



(b)

Picture 6: load distribution of (a)Varus and (b) Valgus of before and after correction results

Discussion

This study used a numerical FEA method to evaluate tibiofemoral load distribution before and

after alignment correction in varus and valgus knees. The aim was to understand how the level of osteotomy (tibia or femur) affects load redistribution under the same modeling conditions. Before correction, varus alignment showed higher stress in the medial compartment, while valgus alignment showed higher stress in the lateral compartment. Some points of these findings agree with previous studies (Chantarapanich et al., 2009; Wu et al., 2022). Biomechanically, this happens because changes in the HKA angle shift the mechanical axis and change the joint moment at the knee. In neutral alignment, the load passes near the center of the knee, allowing balanced load sharing. In varus alignment, the axis shifts medially, increasing the adduction moment and medial compressive force. In valgus alignment, the axis shifts laterally, increasing the abduction moment and lateral compressive force (Chantarapanich et al., 2009; Li et al., 2020).

After correction, the improvement in load distribution depended on both the correction angle and the level of deformity. In varus knees, (MOWHTO) provided the greatest reduction in medial stress. Correction to (0° HKA) reduced medial loading, and mild valgus overcorrection (+5° HKA) further improved load balance, like previous findings (Liu et al., 2019; Pan et al., 2022). DFO also reduced medial stress in varus knees but was less effective at (0° HKA), likely because most varus deformities come from the tibia (Martay et al., 2018). In valgus knees, femoral-based correction, especially LOWDFO, more effectively reduced lateral overload and restored better balance between compartments. Tibial-based correction increased lateral stress, supporting recommendations that femoral osteotomy is more suitable for valgus deformity (Ismailidis et al., 2023).

However, it is important to distinguish between what is directly demonstrated by the model and what can be interpreted as a potential clinical implication. In the present study, some corrected valgus cases showed only minimal changes in medial–lateral load ratios compared to the pre-correction condition. This indicates that correction to 0° HKA does not necessarily result in substantial load redistribution. Biomechanically, this occurs because the correction angle is insufficient to shift the mechanical axis adequately toward the medial compartment, and the weight-bearing line remains relatively close to the lateral side, maintaining lateral load dominance (Wu et al., 2022; Li et al., 2022). Therefore, while alignment correction toward neutral reduces excessive lateral loading, it does not always achieve effective redistribution. In contrast, slight varus overcorrection shifts the mechanical axis further toward the medial compartment, allowing a more effective transfer of load away from the lateral side and resulting in a more favorable distribution pattern. This is consistent with previous findings, where slight varus alignment produced more effective stress redistribution in valgus knees compared to neutral alignment (Wu et al., 2022). However, excessive varus correction should be avoided,

as it may increase medial compartment loading and risk degeneration.

From a surgical point of view, HTO is performed to correct varus deformity and reduce medial compartment overload. The two main techniques are MOW and LCW. LCW provides strong stability and allows earlier weight-bearing, but it removes bone, may shorten the limb, and carries a higher risk of nerve injury. MOW preserves bone and allows more precise correction of the mechanical axis, and makes future knee replacement easier, although it may have a slightly higher risk of nonunion or tibial slope changes. Between two techniques, MOWHTO is increasingly preferred because it preserves bone and provides better control of alignment, while LCW is valued for its stability. HTO is generally preferred for varus deformity. (Li et al., 2021). DFO is used to correct valgus malalignment and reduce lateral compartment overload. The two main techniques are LOW and MCW. LOWWDFO allows gradual and precise correction and preserves bone stock, but it is less stable and has a higher risk of hinge fracture or delayed healing. In contrast, MCWDFO removes a bone wedge and compresses the osteotomy site, providing greater mechanical stability, higher stiffness, and more reliable bone healing. Overall, MCWDFO is often preferred due to its superior stability and predictable outcomes, while LOWDFO is selected in specific cases depending on deformity type and surgical (Ferrera & Menetrey, 2022). These findings highlight the importance of anatomy-based surgical decision-making. Osteotomy should be selected according to the origin of the deformity and its mechanical effect: tibial-based correction (especially MOWHTO) is more effective for varus knees, while femoral-based correction (MCWDFO or LOWDFO) is more appropriate for valgus knees. Although the ideal correction angle remains debated, mild overcorrection may improve load redistribution. Overall, alignment-guided planning based on biomechanical principles can optimize postoperative load balance and potentially reduce long-term joint degeneration.

Some limitations of this study should be recognized. The analysis was performed under static loading conditions. In addition, simplified material properties were used in the knee model, as cartilage, menisci, and ligaments were represented using linear elastic, homogeneous, and isotropic material properties without subject-specific ligament characteristics. Furthermore, only correction angles of (0° HKA and +5° HKA) were evaluated. The application of more advanced material models, such as nonlinear or hyperelastic properties for soft tissues, and inclusion of dynamic loading conditions may provide more precise biomechanical results and should be addressed in future investigations.

Conclusion

This study evaluated tibiofemoral load distribution in varus and valgus knees before and after

alignment correction using FEA. The results showed that load distribution is strongly dependent on the HKA angle. Varus alignment increased medial compartment loading, while valgus alignment increased lateral compartment loading. For varus knees, tibial-based correction especially (MOWHTO) showed the greatest reduction in medial loading. Correction to neutral alignment (0° HKA) reduced medial load, and mild valgus overcorrection (+5° HKA) further improved load redistribution. For valgus knees, femoral-based correction particularly (LOWDFO) more effectively restored mediolateral load balance by reducing excessive lateral compartment loading. Tibial-based correction in valgus alignment was less favorable in terms of load redistribution. Overall, the findings suggest that optimal load distribution depends on both the correction HKA angle and the anatomical level of osteotomy. Tibial-based osteotomy may be more suitable for varus deformity, while femoral-based osteotomy may be more appropriate for valgus deformity under static loading conditions.

Overall, FEA is a reliable numerical method for evaluating knee biomechanics; however, it provides approximate results rather than absolute accuracy. The findings of this study demonstrate that tibiofemoral load distribution varies with HKA alignment, showing clear tendencies of load transfer between the medial and lateral compartments. Despite the use of a simplified model and underlying assumptions, including limited mechanical validation, the results remain consistent under controlled conditions. Furthermore, the observed outcomes are primarily influenced by alignment correction and input parameters rather than the specific knee model used. Therefore, this approach is suitable for analyzing biomechanical tendencies and enabling comparative evaluation of load distribution changes.

Recommendations

Future studies should consider dynamic loading conditions and muscle forces to better represent physiological knee mechanics. The use of nonlinear and viscoelastic material models, subject-specific ligament properties, and a wider range of correction angles would improve physiological accuracy and enhance the clinical applicability of numerical models for individualized osteotomy planning. The mechanical links between postoperative alignment and complications including plate failure, cartilage degeneration, and nonunion are still not clearly defined, and a standardized "safe correction angle" has yet to be established.

Reference

Asgaonkar, B., & Ghugare, N. T. (2020). The relationship between emotional status, pain severity, radiographic osteoarthritis, and quality of life in patients with knee osteoarthritis. *Indian*

Journal of Physiotherapy and Occupational Therapy, 14(1), 55–60.

Bergstein, V. E., Weinblatt, A. I., Taylor, W. L., IV, & Long, W. J. (2024). Total knee arthroplasty survivorship and outcomes in young patients: A review of the literature and 40-year update to a longitudinal study. *Archives of Orthopaedic and Trauma Surgery, 144*, 4077–4083.

Carranza, V. A., Getgood, A., Willing, R., & Burkhart, T. A. (2025). Comparing simulation approaches used in finite element modelling of a medial opening wedge high tibial osteotomy. *Annals of Biomedical Engineering, 53*(4), 731–739.

Chantarapanich, N. (2010). *Determination of the correction angle for the treatment of early stage knee osteoarthritis by high tibial osteotomy* (Master's thesis). Sirindhorn International Institute of Technology, Thammasat University, Thailand.

Chantarapanich, N., Nanakorn, P., Chernchujit, B., & Sitthiseripratip, K. (2009). A finite element study of stress distributions in normal and osteoarthritic knee joints. *Journal of the Medical Association of Thailand, 92*(Suppl. 6), S97–S103.

Chen, J., Li, J., Zhang, H., Feng, W., Ye, P., Qi, X., Deng, P., Li, Y., Huang, Y., Zeng, J., & Zeng, Y. (2024). Bone healing and clinical outcomes following medial opening wedge high tibial osteotomy using wedge-shaped cancellous allograft. *Orthopaedic Surgery, 16*(1), 86–93.

Cui, A., Li, H., Wang, D., Zhong, J., Chen, Y., & Lu, H. (2020). Global, regional prevalence, incidence, and risk factors of knee osteoarthritis in population-based studies. *EClinicalMedicine, 29-30*, 100587.

Dabis, J., Templeton Ward, O., Lacey, A. E., Narayan, B., & Trompeter, A. (2017). The history, evolution, and basic science of osteotomy techniques. *Strategies in Trauma and Limb Reconstruction, 12*(3), 169-180.

Donahue, T. L. H., Hull, M. L., Rashid, M. M., & Jacobs, C. R. (2003). How the stiffness of meniscal attachments and meniscal material properties affect tibiofemoral contact pressure computed using a validated finite element model. *Journal of Biomechanics, 36*(1), 19-34.

Duethman, N. C., Bernard, C. D., Camp, C. L., Krych, A. J., & Stuart, M. J. (2019). Medial closing wedge distal femoral osteotomy. *Clinical Sports Medicine, 38*(3), 361-373.

Elyasi, E., Cavalié, G., Perrier, A., Graff, W., & Payan, Y. (2021). Complications of open wedge high tibial osteotomy: A systematic review from clinical and biomechanical perspectives. *Applied Bionics and Biomechanics, 2021*, 6633891.

Ferrera, A., & Menetrey, J. (2022). Optimizing indications and technique in osteotomies around the knee. *Journal of Experimental Orthopaedics, 9*, 396-403.

Fujisawa, Y., Masuhara, K., & Shiomi, S. (1979). The effect of high tibial osteotomy on osteoarthritis of the knee: An arthroscopic study of 54 knee joints. *Orthopedic Clinics of North America, 10*, 585-608.

Higuchi, T., Koseki, H., Yonekura, A., Imai, C., Tomonaga, I., Sunagawa, S., Matsumura, U., & Osaki, M. (2024). Comparison of short-term clinical outcomes between open-wedge high tibial osteotomy and tibial condylar valgus osteotomy. *BMC Musculoskeletal Disorders, 25*, 98.

Ismailidis, P., Schmid, C., Werner, J., Nüesch, C., Mündermann, A., Pagenstert, G., & Egloff, C. (2023). Distal femoral osteotomy for the valgus knee: Indications, complications, and outcomes. *Archives of Orthopaedic and Trauma Surgery, 143*, 6147-6157.

Kang, K. T., Koh, Y. G., Lee, J. A., Lee, J. J., & Kwon, S. K. (2020). Biomechanical effect of lateral hinge fracture in medial opening wedge high tibial osteotomy: A three element study. *Journal of Orthopaedic Surgery and Research, 15*, 63.

Kang, Y., Kim, J., Sim, J. A., Moon, M., Park, J.-C., Cho, S. H., & Lee, B. H. (2023). Stress effect in the knee joint based on the fibular osteotomy level and varus deformity: A finite element analysis study. *Bioengineering, 10*(9), 1003.

Karkosch, R. F., Schwarze, M., Smith, T., Petri, M., Pastor, M.-F., & Horstmann, H. (2024). Medial open wedge osteotomy yields comparable stability to lateral open wedge procedure on the distal femur. *PLOS ONE, 19*(9), e0310869.

Kellgren, J. H., & Lawrence, J. S. (1957). Radiological assessment of osteoarthrosis. *Annals of the Rheumatic Diseases, 16*(4), 494–502.

Kim, Y.-T., Han, B.-S., Kim, J. B., Sa, J. K., Hong, J. H., Son, Y., Han, J.-H., Do, S., Chae, J. S., & Bae, J.-K. (2024). HKA-Net: Clinically adapted deep learning for automated measurement of hip–knee–ankle angle on lower limb radiography for knee osteoarthritis assessment. *Journal of Orthopaedic Surgery and Research, 19*, 777.

Li, M., Chang, H., Wei, N., Chang, W., Yan, Y., Jin, Z., & Chen, W. (2020). Biomechanical study on the stress distribution of the knee joint after tibial fracture malunion with residual varus–valgus deformity. *Orthopaedic Surgery, 12*(3), 983-989.

Li, O. L., Pritchett, S., Giffin, J. R., & Spouge, A. R. I. (2021). High tibial osteotomy: An update for

radiologists. *American Journal of Roentgenology, 217*(6), 1290-1302.

Liu, X., Chen, Z., Gao, Y., Zhang, J., & Jin, Z. (2019). High tibial osteotomy: Review of techniques and biomechanics. *Journal of Healthcare Engineering, 2019*, 8363128.

Mahaisavariya, B., Sitthiseripratip, K., Tongdee, T., Bohez, E. L. J., Vander Sloten, J., & Oris, P. (2002). Morphological study of the proximal femur: A new method of geometrical assessment using 3-dimensional reverse engineering. *Medical Engineering & Physics, 24*, 617-622.

Martay, J. L. B., Palmer, A. J. R., Bangerter, N. K., Clare, S., Monk, A. P., Brown, C. P., & Price, A. J. (2018). Safe correction zone for high tibial osteotomy: A modelling investigation. *The Knee, 25*(2), 286-295.

Meisterhans, M., Flury, A., Zindel, C., Zimmermann, S. M., Vlachopoulos, L., Snedeker, J. G., & Fucentese, S. F. (2023). Finite element analysis of medial closing and lateral opening wedge distal femoral osteotomies in relation to hinge fractures. *Journal of Experimental Orthopaedics, 10*, 33.

Nakamura, R., Amemiya, M., Matsumoto, K., Yoshida, I., Kawashima, F., Shimakawa, T., & Okano, A. (2024). Bone substitute fracture in open wedge high tibial osteotomy: Comparison of two different bone substitutes. *Journal of Orthopaedic Surgery and Research, 19*, 243.

Osmani, H. T., Gupta, R., Earl, R., Tomaszczyk, S., Turmezei, T., Segal, N. A., Sutcliffe, M., & Melton, J. (2024). Finite element analysis confirms the optimal apex position in medial opening wedge high tibial osteotomy to avoid lateral hinge fracture. *Journal of Experimental Orthopaedics, 11*, e70042.

Pan, C. S., Wang, X., Ding, L. Z., Zhu, X. P., Xu, W. F., & Huang, L. X. (2022). Optimal bone graft position in medial open wedge high tibial osteotomy: A finite element analysis. *Computer Methods and Programs in Biomedicine, 224*, 107253.

Peña, E., Calvo, B., Martínez, M. A., & Doblaré, M. (2006). A three-dimensional finite element analysis of the combined behavior of ligaments and menisci in the healthy human knee joint. *Journal of Biomechanics, 39*(9), 1686–1701.

Trad, Z., Barkaoui, A., Chafra, M., & Tavares, J. M. R. S. (2018). Finite element analysis applications in biomechanical studies of the knee joint. In *FEM analysis of the human knee joint* (pp. 35–60). Springer.

Wu, Y., Jin, X., Zhao, X., Wang, Y., Bai, H., Lu, B., Tong, X., Ma, J., & Ma, X. (2022). Computer-

aided design of distal femoral osteotomy for the valgus knee and effect of correction angle on joint loading by finite element analysis. *Orthopaedic Surgery, 14*(11), 2904–2913.

Wylie, J. D., & Maak, T. G. (2016). Distal femoral osteotomy for the valgus knee: Medial closing wedge versus lateral opening wedge. *Arthroscopy: The Journal of Arthroscopic and Related Surgery, 32*(10), 2141-2147.

Xu, K., Wang, T., Yu, T., Zhao, X., Zhang, Y., & Zhang, L. (2024a). Comparison of different degrees of varus deformity correction with open-wedge high tibial osteotomy: A retrospective study over 5 years. *Journal of Orthopaedic Surgery and Research, 19*, 93.

Xu, K., Zhang, L., Yu, T., Zhao, X., & Zhang, Y. (2024b). Effect of knee valgus deformity on symptomatic venous thromboembolism and prosthesis revision risk after total knee arthroplasty: A multicenter retrospective study. *Orthopaedic Surgery, 16*(2), 500–508.

Zhang, Z., Tao, H., Zhao, Y., Xiang, W., Cao, H., & Tao, F. (2023). High tibial osteotomy improves balance control in patients with knee osteoarthritis and a varus deformity. *Journal of Orthopaedic Surgery and Research, 18*, 538.

SLIDING AND NON-SLIDING BOUNDARY CONDITIONS IN BONE-BORNE MINI SCREW-ASSISTED RAPID PALATAL EXPANSION: A FINITE ELEMENT NUMERICAL ANALYSIS

Viet Hoang[1], Nattapon Chantarapanich[2*], Samroeng Inglam[3], and Anand Marya[4]

[12]Faculty of Engineering at Sriracha, Kasetsart University Sriracha Campus, Chonburi, Thailand
[3]Faculty of Dentistry, Thammasat University, Pathumthani, Thailand
[4]Faculty of Dentistry, University of Puthisastra, Phnom Penh, Cambodia
[*]**Corresponding author: nattapon@eng.src.ku.ac.th**

Abstract

Objective: To demonstrate the influence of sliding and non-sliding boundary condition on the biomechanical behavior of bone-borne mini screw-assisted rapid palatal expansion (MARPE) using three-dimensional finite element analysis, with particular emphasis on stress distribution and mid palatal suture opening effectiveness.

Materials and Methods: A three-dimensional finite element model of a bone-borne MARPE system was constructed based on Cone Beam Computed Tomography (CBCT) data from a 20-year-old male patient. Two numerical models with identical geometry, material properties, loading conditions, and global cranial constraints were developed. The difference between the models was the internal boundary condition of the MARPE system: a non-sliding boundary condition and a sliding boundary condition along the connector. Von Mises stress in the MARPE body and mini screws, major principal stress in cortical and cancellous bone, and mid palatal suture opening at anterior and posterior points were evaluated following 0.16 mm of expander activation.

Results: The two boundary condition models demonstrated markedly different biomechanical responses. The non-sliding model exhibited lower von Mises stress in both the MARPE body and mini screws, as well as reduced major principal stress in cortical bone. However, it produced limited mid palatal suture opening and a pronounced V-shaped expansion pattern. In contrast, the sliding boundary condition resulted in greater anterior suture opening, a higher posterior-to-anterior expansion ratio, and a more parallel expansion pattern, despite increased localized stress concentration at the mini screw-cortical bone interface.

Conclusions: Sliding contact assumptions may better reflect the clinical activation mechanics of bone-borne MARPE by permitting effective lateral force transmission and producing expansion patterns closer to clinical observations. Although non-sliding assumptions reduce stress magnitudes, they may artificially constrain the expander mechanism and underestimate mid palatal suture opening. Therefore, boundary condition selection should be explicitly justified to ensure clinically meaningful finite element predictions in MARPE research.

Keywords: MARPE, Mini Screws, Expansion, Boundary Conditions, Mid Palatal

Introduction

Mini Screw-Assisted Rapid Palatal Expansion (MARPE) has been widely used for the treatment of maxillary transverse deficiency, posterior crossbite or buccal corridor treatment to improve smile aesthetic, particularly in late adolescents and adults, as it increases skeletal expansion and reduces

dentoalveolar side effects compared with conventional tooth-borne expanders. (Choi et al., 2016; Brunetto et al., 2017) MARPE was invented in 2010 by professor Kee Joon Lee who used tooth-bone borne design (Lee et al., 2010) and after that, a lot of studies have demonstrated that bone-borne MARPE systems can generate effective lateral expansion forces through mini screws directly anchored in the maxillary bone. (Lin et al., 2015) Finite element (FE) analysis has been widely applied to check the biomechanical behavior of MARPE, including maxillary displacement, stress distribution, and mid-palatal suture response. However, it is well recognized that the predictive accuracy of FE strongly depends on modeling assumptions, particularly the definition of boundary conditions (Jafari et al., 2003). Previous orthodontic FE studies have shown that different boundary condition formulations, such as bonded, tied, frictionless, or frictional contact can substantially alter predicted force transmission and deformation patterns (Jafari et al., 2003; Gautam et al., 2007; Rivera-Tapia et al., 2025). Despite the acknowledged importance of boundary conditions, some papers focused on and indicated boundary conditions is one of important methodology in FE analysis (Rivera-Tapia et al., 2025). Nevertheless, the way expansion forces are generated by the MARPE device itself and transmitted laterally under different boundary condition assumptions has not been explicitly addressed. In particular, the effect of sliding versus non-sliding boundary conditions on the lateral force generation and transmission mechanism of bone-borne MARPE remains unconclusive.

Research Objectives

1. The objective of this study was to assess the effect of sliding and non-sliding boundary condition assumptions on the biomechanical predictions of a FE model of bone-borne MARPE.
2. To evaluate whether sliding boundary conditions provide a more clinically relevant representation of bone-borne MARPE mechanics than non-sliding boundary conditions.

Literature Review

FE analysis in orthodontics, particularly in studies involving MARPE, is highly dependent on the definition of boundary conditions, as these directly govern force transmission, stress distribution, and deformation patterns within numerical models. Accordingly, methodological guidelines emphasize the need for clear and transparent reporting of boundary condition strategies to ensure reproducibility and meaningful interpretation of results. Previous studies have highlighted that appropriate specification of boundary conditions and loading configurations represents a fundamental step in FE modeling; furthermore, improving the clinical applicability of orthodontic FE analysis requires detailed anatomical reconstruction, physiologically relevant boundary conditions, and validation against longitudinal CBCT data (Rivera-Tapia et al., 2025; Singh et al., 2016). Such methodological advancements are essential for the development of evidence-based, patient-specific MARPE protocols and for improving treatment planning, safety, and long-term stability. Most existing MARPE FE studies have adopted conservative boundary condition formulations focused primarily on global cranial stability, typically by fixing nodes at the foramen magnum in all translational and rotational directions (Mamboleo et al., 2024; Sucu et al., 2021; MacGinnis et al., 2014)

While this approach effectively prevents rigid-body motion of the craniofacial model, these studies have predominantly emphasized outcome-based parameters, such as stress distribution and skeletal displacement. In contrast, the underlying mechanical mechanisms by which expansion forces are generated and transmitted through the MARPE appliance itself have received comparatively limited attention.

From a modeling perspective, boundary conditions in orthodontic FE analyses can be broadly categorized into global cranial constraints and internal contact definitions within the appliance-bone system. Although cranial fixation strategies are widely reported in the MARPE literature, internal

boundary conditions governing appliance activation mechanics, such as relative motion between expander components have rarely been systematically investigated. Consequently, a clear gap exists in literature, as no previous studies have directly compared sliding and non-sliding boundary condition assumptions with respect to lateral expansion force transmission in bone-borne MARPE designs.

Research Methodology

1. Study design

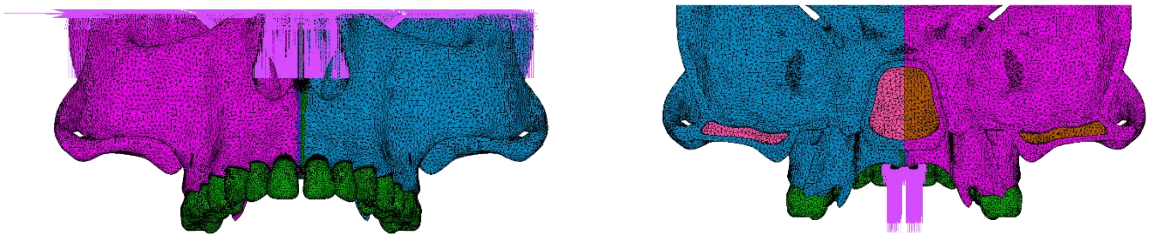
A three-dimensional FE analysis was conducted to compare sliding versus non-sliding boundary condition assumptions on lateral expansion force transmission in a bone-borne MARPE design. Two numerical models were established that were identical in geometry, material properties, mesh density, loading protocol, and global constraints; the only difference between models was the boundary condition definition governing relative motion within the MARPE assembly (sliding vs non-sliding). Craniofacial geometry of the maxilla and adjacent cranial structures was reconstructed from CBCT data of male patient, which were segmented using Blue Sky Plan (Blue Sky Bio Digital, US) to generate Stereolithography (STL) files representing the cortical bone, cancellous bone, and mid palatal region. The STL models were subsequently imported into CAD software (VISI, Hexagon AB, Sweden) for surface refinement, solid modeling, and integration with the MARPE appliance geometry. A bone-borne MARPE Sel·P appliance and 4 mini screws designed from Professor Kee Joon Lee and manufactured by Myungsung Company (South Korea) was modeled in CAD software and positioned on the palate according to the clinical design in the middle position. The expander body and connecting body were defined as stainless steel. Four mini screws (1.8 mm × 12 mm) were modeled and inserted in the middle palatal position, corresponding clinically to the interradicular level between the maxillary second premolar and first molar on both sides, following a symmetrical configuration. The screws insertion through 4 holes of MARPE with depth and engagement length were standardized across models. MSC Marc®, MSC software was used for the FE analysis. The von Mises stress distributions after 0.16 mm of MARPE and Mini screws were evaluated as 2 cases with different boundary conditions.

All materials were assumed to be homogeneous, isotropic, and linearly elastic. Elastic properties were assigned based on commonly used craniofacial FE analysis values. Representative values include:

- Cortical bone: E = 15,000 MPa, ν = 0.30
- Cancellous bone: E = 1,000 MPa, ν = 0.30
- MARPE body (Stainless Steel): E □ 200,000 MPa, ν = 0.30
- Tooth (Dentin): E □ 20,000 MPa, ν = 0.30
- Titanium alloy mini screws (Ti): E □ 110,000 MPa, ν = 0.34

To prevent rigid-body motion of the skull model and ensure numerical stability, nodes at the foramen magnum region were determined as the boundary condition, and all the displacements were restricted to this area, the 2 parts of the maxilla were separated so that they could move with expansion forces laterally with respect to the vertical plane of symmetry. The 2 boundary-conditioning scenarios were defined:

- Model A – Non-sliding boundary condition with only 1 boundary condition at the foramen magnum, as shown in Figure 1(a)
- Model B – Sliding boundary condition was designed same with model A but with 2 boundary conditions, 1st at foramen magnum same as Model A and 2nd at the MARPE body with lateral side, as shown in Figure 1(b)



(a) (b)

Picture 1: (a) Model A with 1st boundary condition, and (b) Model B with the 2nd boundary conditions

Results

Two FE models representing sliding and non-sliding boundary condition scenarios were evaluated. Outcome measures included the von Mises stress distribution within the MARPE appliance and the four mini screws, major principal stress of cortical and cancellous bone, as well as the magnitude of mid palatal suture opening measured at the anterior point and posterior point to compare the effectiveness of two techniques and V shape opening result. The two boundary condition scenarios, sliding and non-sliding, produced markedly different biomechanical responses in terms of expansion effectiveness, stress distribution, and mid palatal suture opening.

1. Comparison of Von Mises Stress in MARPE appliance, mini screws and major principal stress between 2 boundaries condition.

1.1 Von Mises Stress

The sliding and non-sliding boundary condition models produced distinct von Mises stress profiles in both the MARPE body and mini screws. Under the sliding boundary conditions, the von Mises stress within the MARPE body was 728.56 MPa, whereas the peak von Mises stress among the four mini screws reached 1598.64 MPa. In contrast, under the non-sliding boundary conditions, the MARPE body exhibited a markedly lower von Mises stress of 253.55 MPa, and the mini screws showed a reduced peak stress of 910.45 MPa.

Overall, when considering von Mises stress magnitude alone as an indicator of mechanical safety, the non-sliding boundary condition demonstrated a more favorable state of stress, with substantially lower stresses in both the MARPE body (253.55 MPa vs. 728.56 MPa) and the mini screws (910.45 MPa vs. 1598.64 MPa) compared with the sliding condition. These results indicate that allowing sliding at the defined interfaces led to increased stress concentration at the mini screw level and higher stress within the appliance body. The result was shown in Picture 2.

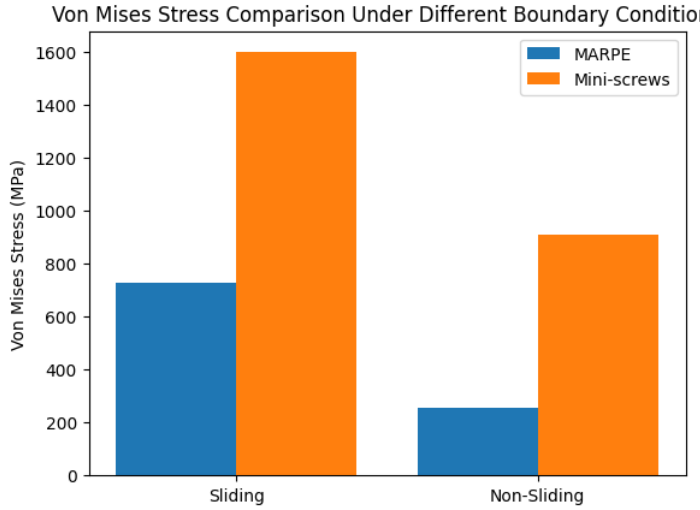
1.2 Major Principal Stress Result

Major principal stress in the surrounding bone also differed between the two boundary condition scenarios. In the sliding model, the maximum major principal stress in the cortical bone reached 221.13 MPa, almost twice the value observed in the non-sliding model (112.24 MPa). In contrast, stress levels in the cancellous bone were low and similar in both models, measuring 8.31 MPa for sliding and 8.64 MPa for non-sliding. These results suggest that allowing sliding promotes greater stress transfer to the cortical bone, while cancellous bone stress is relatively unaffected by the choice of boundary condition. The result was shown in Picture 3.

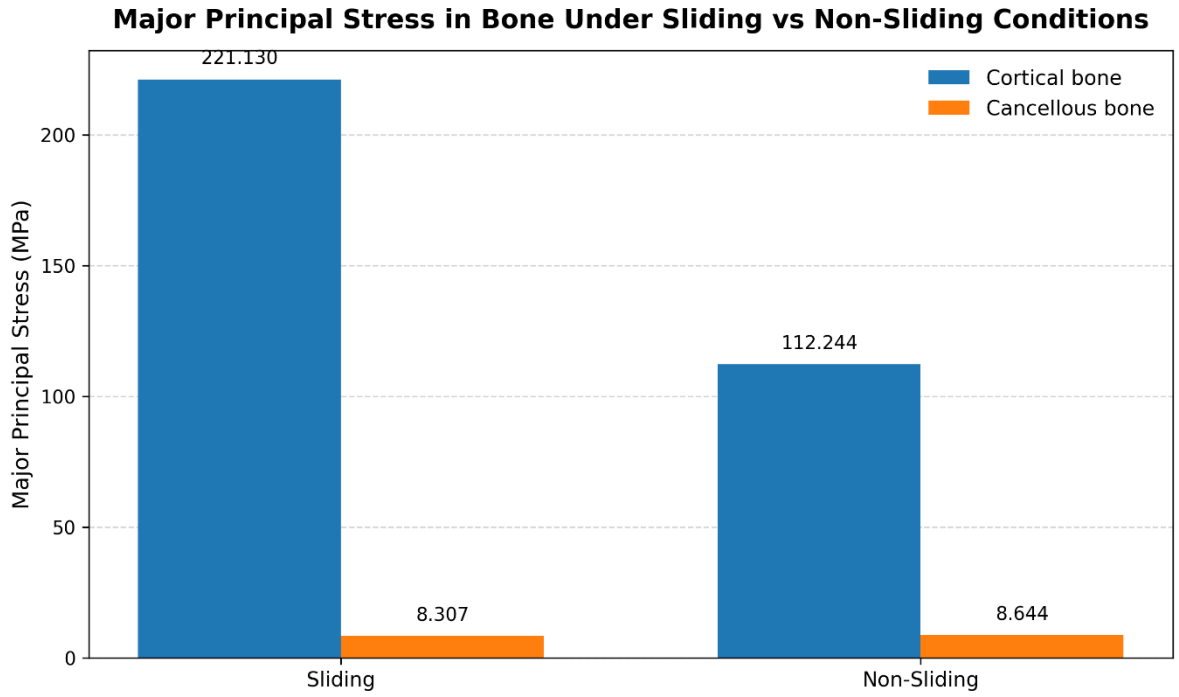
2. Effectiveness of mid palatal suture opening effectiveness between 2 boundaries condition

2.1 Mid palatal suture opening effectiveness

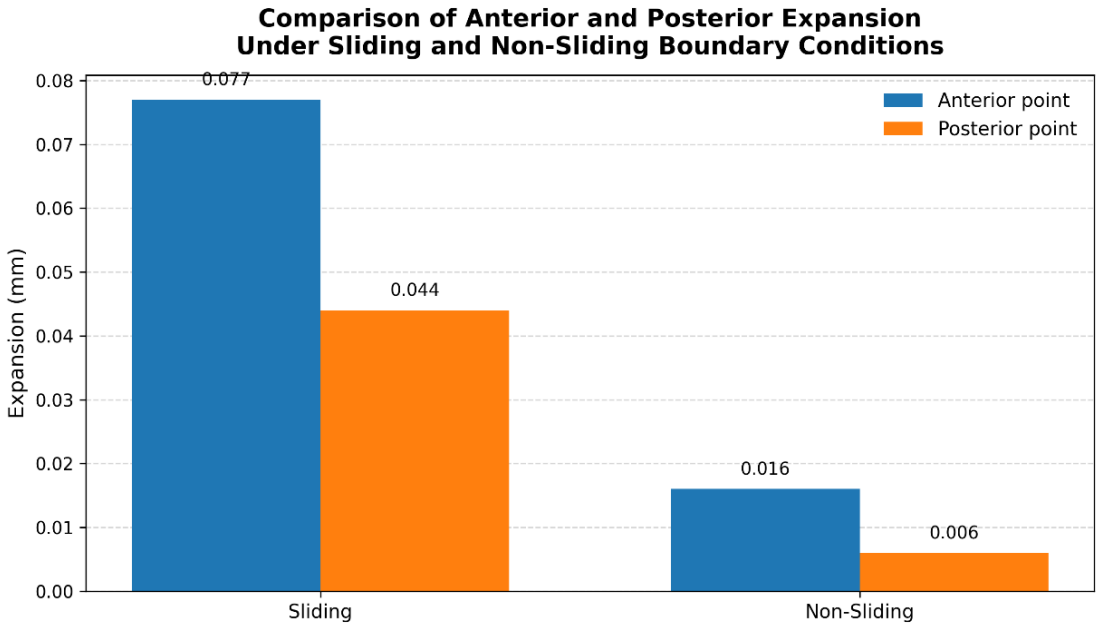
The two boundary condition models showed differences in both the amount and pattern of mid palatal suture opening. In the non-sliding condition, the opening at the anterior point was limited to 0.016 mm and posterior point is 0.006 mm. In contrast, the sliding condition produced a larger anterior opening of 0.077 mm and posterior point 0.044 mm, corresponding to more than a twofold increase. Both models exhibited a V-shaped expansion pattern, with greater opening anteriorly than posteriorly; however, the degree of parallelism differed. The posterior-to-anterior expansion ratio was higher in the sliding model (0.044/0.077 ≈ 0.57) than in the non-sliding model (0.006/0.016 ≈ 0.38), indicating a less V shape and a more parallel expansion pattern when sliding was permitted. The result was shown in Picture 4 and 5.



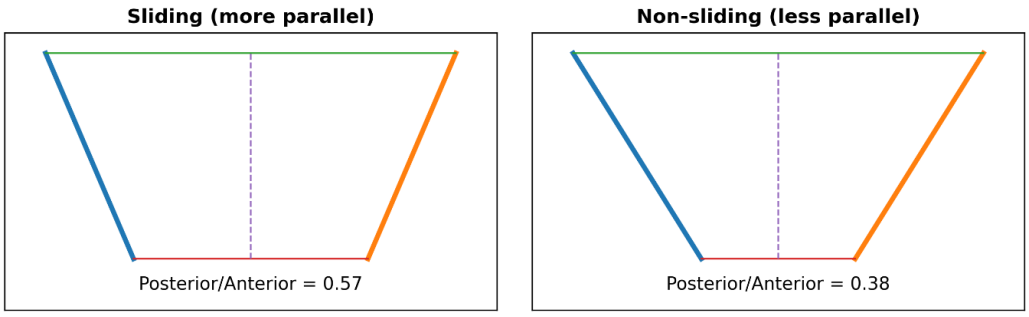
Picture 2: Comparison of von Mises stress distribution in the MARPE body and mini screws under sliding and non-sliding boundary conditions



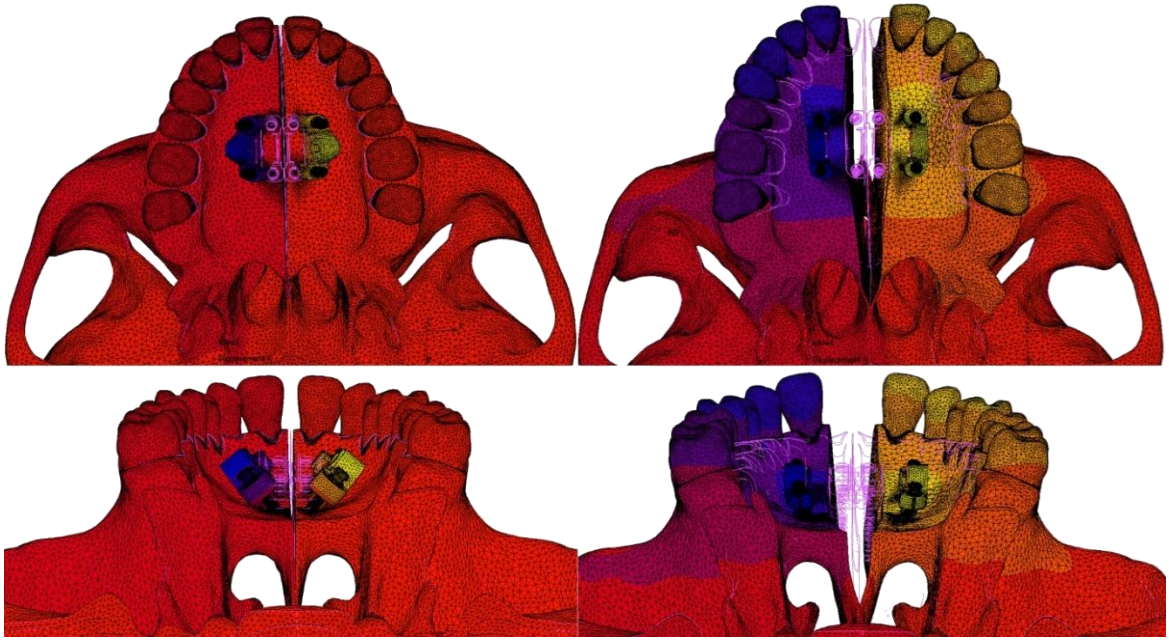
Picture 3: Comparison of major principal stress in cortical and cancellous bone under sliding and non-sliding boundary conditions



Picture 4: Comparison of anterior and posterior suture opening result under sliding and non-sliding boundary conditions



Picture 5: Comparison of V-shaped expansion patterns under the two boundary conditions



Picture 6: Comparison of the two boundary condition techniques. Under the non-sliding boundary condition (left), the two parts of the MARPE appliance separated non-parallel pattern after activation. In contrast, the sliding boundary condition (right) allows a more parallel lateral displacement of the MARPE parts, which more closely reflects the expected clinical behavior of bone-borne MARPE.

Discussion

The present FE analysis demonstrates that boundary condition is effective and one of the most important aspects in determining the predicted biomechanics in orthodontics and especially in bone-borne MARPE (MacGinnis et al., 2014; Knop et al., 2015). In FE modeling, boundary conditions define how the numerical model interacts with its environment, constrain motion, and regulate force transmission pathways. Consequently, even subtle variations in boundary condition assumptions can substantially influence predicted stress distribution, displacement patterns, and overall biomechanical behavior. In orthodontics, where clinical decisions often rely on interpreting stress concentration and skeletal response patterns, inaccurate or oversimplified boundary condition formulations may lead to misleading conclusions regarding appliance performance and treatment efficiency. Although previous MARPE FE studies have primarily focused on stress distribution and skeletal displacement outcomes,

the influence of internal boundary conditions governing relative motion within the expander system has received limited attention (Jafari et al., 2003; Gautam et al., 2007; Rivera-Tapia et al., 2025). The findings of this study indicate that sliding versus non-sliding boundary conditions fundamentally alter both expansion effectiveness and stress distribution.

In the results, although the stress distribution of non-sliding boundary condition shown the stress distribution in MARPE, mini screws, and cortical bone are less than the sliding boundary condition. However, the non-sliding condition resulted in limited suture opening and a pronounced V-shaped expansion pattern, indicating reduced expansion efficiency. This may not correlate to actual clinical observation. As mentioned earlier, boundary conditions play a key role in FE analysis, as they strongly influence how closely the simulated results resemble real clinical behavior. Previous study has shown that inappropriate boundary condition assumptions can lead to results that fail to reflect the true mechanical response of appliances (Hu et al, 2017). In our study, this limitation is clearly illustrated in Picture 6, where different boundary condition assumptions produce markedly different biomechanical outcomes. In most MARPE FE studies, boundary conditions are restricted to global cranial fixation at the foramen magnum, which may oversimplify appliance activation mechanics and lead to expansion patterns that deviate from clinical observations. Under such circumstances, the MARPE system is not similar with in-vivo activation of MARPE (Thi Hong Thuy et al., 2025; Hoang et al., 2025b) In contrast, when an internal lateral sliding boundary condition was introduced, the simulated expansion pattern more closely resembled clinical behavior. These findings emphasize that MARPE FE models should incorporate clinically relevant internal contact definitions, rather than relying solely on global cranial constraints, to accurately simulate lateral force transmission and appliance activation (Hoang et al., 2025a; Hong., 2025b). Without proper boundary conditions, it may mislead dental surgeons in term of clinical utilization, underscoring the need for careful and deliberate boundary condition selection in orthodontic FE analysis.

Conclusion

Sliding boundary condition better represents the actual clinical behavior of bone-borne MARPE and should be preferred in FE simulations. Non-sliding assumptions may restrict lateral force transmission, may not accurately reflect clinical reality, so lead to inaccurate predictions of skeletal expansion. Therefore, researchers should carefully select and justify boundary condition definitions, as inappropriate assumptions can compromise the validity and clinical relevance of numerical outcomes.

References

Brunetto, D. P., Sant'Anna, E. F., Machado, A. W., & Moon, W. (2017). Non-surgical treatment of transverse deficiency in adults using microimplant-assisted rapid palatal expansion (MARPE). *Dental Press Journal of Orthodontics, 22*(1), 110-125. https://doi.org/10.1590/2177-6709.22.1.110-125.sar

Choi, S. H., Shi, K. K., Cha, J. Y., Park, Y. C., & Lee, K. J. (2016). Nonsurgical miniscrew-assisted rapid maxillary expansion results in acceptable stability in young adults. *The Angle Orthodontist, 86*(5), 713–720. https://doi.org/10.2319/101415-689.1

Gautam, P., Valiathan, A., & Adhikari, R. (2007). Stress and displacement patterns in the craniofacial skeleton with rapid maxillary expansion: A finite element method study. *American Journal of Orthodontics and Dentofacial Orthopedics, 132*(5), e1–e11. https://doi.org/10.1016/j.ajodo.2006.09.044

Hoang, V., Marya, A., d'Apuzzo, F., & Nucci, L. (2025a). The clinical applications and outcomes of digital MARPE in orthodontics: A scoping review. *Seminars in Orthodontics, 31*(2), 299-309. https://doi.org/10.1053/j.sodo.2024.10.002

Hoang, V., Tran, P. H., & Dang, T. T. (2025b). Buccal corridor and gummy smile treatment with MARPE and gingivoplasty: A 2-year follow-up case report. *APOS Trends in Orthodontics, 15*, 352–359. https://doi.org/10.25259/APOS_216_2023

Hu, P., Wu, T., Wang, H. Z., Qi, X. Z., Yao, J., Cheng, X. D., Chen, W., & Zhang, Y. Z. (2017). Influence of different boundary conditions in finite element analysis on pelvic biomechanical load transmission. *Orthopaedic Surgery, 9*(1), 115-122. https://doi.org/10.1111/os.12315

Jafari, A., Shetty, K. S., & Kumar, M. (2003). Study of stress distribution and displacement of various craniofacial structures following application of transverse orthopedic forces: A three-dimensional finite element method study. *The Angle Orthodontist, 73*(1), 12-20. https://doi.org/10.1043/0003-3219(2003)073<0012:SOSDAD>2.0.CO;2

Knop, L., Gandini, L. G., Jr., Shintcovsk, R. L., & Gandini, M. R. (2015). Scientific use of the finite element method in orthodontics. *Dental Press Journal of Orthodontics, 20*(2), 119-125. https://doi.org/10.1590/ 2176-9451.20.2.119-125.sar

Lee, K. J., Park, Y. C., Park, J. Y., & Hwang, W. S. (2010). Miniscrew-assisted nonsurgical palatal expansion before orthognathic surgery for a patient with severe mandibular prognathism. *American Journal of Orthodontics and Dentofacial Orthopedics, 137*(6), 830-839. https://doi.org/10.1016/j.ajodo. 2007.10.065

Lin, L., Ahn, H. W., Kim, S. J., Moon, S. C., Kim, S. H., & Nelson, G. (2015). Tooth-borne versus bone-borne rapid maxillary expanders in late adolescence. *The Angle Orthodontist, 85*(2), 253-262. https://doi.org/10.2319/030514-156.1

MacGinnis, M., Chu, H., Youssef, G., Wu, K. W., Machado, A. W., & Moon, W. (2014). The effects of micro-implant assisted rapid palatal expansion (MARPE) on the nasomaxillary complex: A finite element method analysis. *Progress in Orthodontics, 15*(1), 52. https://doi.org/10.1186/ s40510-014-0052-y

Mamboleo, E., Ouldyerou, A., Alsharif, K., Ngan, P., Merdji, A., Roy, S., & Mukdadi, O. M. (2024). Biomechanical analysis of orthodontic miniscrew-assisted rapid palatal expansion on dental and bone tissues: A finite-element study. *Journal of Medical Diagnostics, 7*(4), 041007. https://doi.org/ 10.1115/1.4065589

Rivera-Tapia, E. D., Avila-Vega, C., Lopez-Fajardo, G., Tapia-Bravo, V., Ilbay-Yupa, D., & Mantilla, P. (2025). Assessing the clinical relevance of finite element models in MARPE-induced behaviour of craniofacial structures: A biomechanical review. *Journal of Biomechanics, 193*, 113010. https://doi.org/ 10.1016/j.jbiomech.2025.113010

Singh, J. R., Kambalyal, P., Jain, M., & Khandelwal, P. (2016). Revolution in orthodontics: Finite element analysis. *Journal of International Society of Preventive & Community Dentistry, 6*(2), 110-114. https://doi.org/10.4103/2231-0762.178743

Sucu, M., Yilmaz, B., & Ramoğlu, S. İ. (2021). Evaluation of initial stress distribution and displacement pattern of craniofacial structures with three different rapid maxillary expansion appliance models: A three-dimensional finite element analysis. *Turkish Journal of Orthodontics, 34*(1), 18–25. https://doi.org/10.5152/TurkJOrthod.2021.20006

Thi Hong Thuy, P., Thu Trang, P., Thi Thu Hang, P., & Hoang, V. (2025). Clinical and cone-beam computed tomography outcomes of miniscrew-assisted rapid palatal expansion in the treatment of maxillary transverse deficiency: A prospective study. *Medicine, 104*(38), e44684. https://doi.org/10.1097/MD.0000000000044684

COMPARATIVE ANALYSIS OF DEEP LEARNING MODELS FOR DENTAL AGE ESTIMATION: CEPHALOMETRIC VS. PANORAMIC RADIOGRAPHIC MODALITIES

Sunattanan Petchpan[1], Sujin Wanchat[2*], Nattapon Chantarapanich[3], Anand Marya[4], Na Chihan[5], and Prasitthichai Naronglerdrit[6]

[123]Faculty of Engineering at Sriracha, Kasetsart University Sriracha Campus, Chonburi, Thailand
[456]Faculty of Dentistry, University of Puthisastra, Phnom Penh, Cambodia
[*]**Corresponding Author, E-mail**: sujin@eng.src.ku.ac.th

Abstract

Dental age estimation from radiographic images is important for treatment planning and forensic odontology. This study aimed to develop and evaluate artificial intelligence models using deep-learning-based transfer learning to compare model performance across two radiographic modalities and investigate appropriate model selection. A total of 924 images were collected from a single healthcare facility, comprising 453 cephalometric X-ray images and 471 panoramic X-ray images, covering an age range of 11–45 years, with anonymized file codes assigned. Data preparation and labeling were performed using Roboflow, with images resized to 640×640. Both ResNet101 and EfficientNetB0 models, initialized with ImageNet pre-trained weights (Adam optimizer, learning rate 0.0001, batch size 16), were trained and tested on Google Colab under 16 conditions (image type × augmentation × epochs 10/30 × model), and performance was evaluated using MAE and RMSE for regression, and accuracy, precision, recall, and F1-score for classification (7 age classes in 5-year bins). The results showed that panoramic X-rays exhibited lower errors than cephalometric X-rays. Panoramic X-rays achieved an MAE of approximately 3.03–3.32 years and an RMSE of 4.18–4.29 years, whereas cephalometric X-rays achieved an MAE of approximately 3.77–4.05 years and an RMSE of 5.47–5.51 years. The classification accuracy remained constant across all conditions (0.4038 for cephalometric; 0.5000 for panoramic). Although test accuracy remained unchanged across conditions, error-based metrics revealed differences in prediction performance, indicating that classification accuracy alone may not fully reflect model behavior. This suggests that regression-based evaluation provides a more meaningful basis for model comparison. Increasing the number of training epochs from 10 to 30 tended to reduce MAE. The best cephalometric result was obtained using EfficientNetB0 at 30 epochs with augmentation (MAE 3.77 years; RMSE 5.50 years), while the best panoramic result was obtained using ResNet101 at 30 epochs without augmentation (MAE 3.03 years; RMSE 4.25 years). The proposed MAE threshold of 2.5 years for cephalometric radiographs (Hypothesis H1) was not achieved, establishing a benchmark for future work. Overall, ResNet101 and EfficientNetB0 demonstrated comparable performance, with the optimal model differing by imaging modality. The findings suggest that panoramic radiographs are currently the more suitable modality for deep-learning-based dental age estimation, while cephalometric radiographs may require further optimization before broader clinical or forensic application.

Keywords: Dental Age Estimation, Deep Learning, Cephalometric Radiographs, Panoramic Radiographs, ResNet101, EfficientNetB0

Introduction

Dental age estimation is important for dentists when planning treatments related to craniofacial and jaw growth, including dental and orthodontic treatment planning. Inaccurate dental age assessment poses a risk of serious adverse effects on oral health, particularly when dental age is underestimated. This may affect the timing and appropriateness of treatment planning, which in some cases could

contribute to delayed intervention and poorer oral-health outcomes. For example, untreated dental caries can progress to the dental pulp and to complications affecting surrounding tissues and organs, such as tooth infection or swelling of the face and neck. If left untreated, this may result in tooth loss and permanent consequences for patient management, including identity verification in legal or investigative cases (Aljameel et al., 2023). Because teeth differ across developmental stages, they are used for age estimation. Teeth are therefore a reliable source for age analysis and offer high accuracy, as they are less affected by environmental conditions than skeletal structures and can remain preserved for the longest period (Balel et al., 2025).

Dental age estimation using panoramic radiographs (panoramic images) relies on a frontal dental image that enables visualization of all teeth within the oral cavity. Panoramic radiographs are commonly used in dental practice and are widely adopted in research on age estimation using artificial intelligence models (Mohamed et al., 2023; Singh & Raza, 2022). Deep Convolutional Neural Networks (DCNNs) have played an important role in medical radiographic image analysis (Brahmi et al., 2024) and in improving age assessment accuracy (Lee et al., 2024). Panoramic radiographs are used as a fundamental imaging modality in diagnosis and forensic dental practice. Panoramic dental imaging is a radiographic technique that provides a comprehensive view of the teeth and bilateral oral structures in a single projection, displaying tooth positions, individual teeth, and the jawbone. This allows dentists to evaluate dental conditions and related structures quickly and conveniently, without additional radiation exposure, as it is already part of routine diagnosis (Franco et al., 2024). However, although dental age prediction from panoramic radiographs has been widely applied in combination with Deep Learning (DL) techniques (Koch et al., 2025), limitations remain due to lower image resolution compared with conventional radiographs, which can hinder clear visualization of fine dental details. In addition, panoramic radiographs have viewing-angle limitations that may prevent the complete detection of certain dental or oral conditions. For example, issues located within the tooth or complex dental structures may not be identified accurately. There are also considerations regarding radiation exposure that may affect patient health if performed frequently. Understanding these issues is therefore important for dental diagnosis and treatment (Kim et al., 2021).

Accordingly, further development has been explored using cephalometric radiographs (cephalometric images), which are lateral-view images. A key advantage is that they enable precise measurement and analysis of relationships among anatomical structures within the human head. Two-dimensional radiography provides frontal and lateral skull images, allowing dentists to observe the occlusal relationship between the upper and lower teeth, which is important information for dental age prediction. Previous studies related to age prediction using cephalometric radiographs have primarily focused on skeletal age or bone age and on anatomical landmark analysis (Oztekin et al., 2023), rather than dental age estimation. In addition, although some studies have investigated the use of artificial intelligence models with cephalometric radiographs (Singh & Raza, 2022), existing evidence remains limited, and data-related issues persist. These include the use of small or imbalanced datasets, the lack of cross-modality comparisons, and incomplete reporting of quantitative error measures such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), accuracy, and F1-score (Mohammad et al., 2024).

Khanagar et al. (2021) conducted a study applying artificial intelligence techniques, including deep learning methods and advanced classification algorithms, to cephalometric radiograph analysis to improve performance. Their findings demonstrated meaningful benefits, with a significant increase in the accuracy of dental age detection and prediction using the EfficientNet-B0 model, achieving an RMSE of 1.18 years. In addition, the ResNet-50 model applied to panoramic radiographs achieved an MAE of 1.42 years (Juneja et al., 2021). Moreover, YOLO has been used for analysis and detection in conjunction with a regression model (Oztekin et al., 2023) to improve the identification of structural

features prior to age prediction (Balel et al., 2025). This is particularly important for comparative analysis across models (Juneja et al., 2021) and has therefore been systematically adopted to support further knowledge development in this field (Aljameel et al., 2023).

Therefore, this study aims to develop a method for dental age prediction using deep-learning-based artificial intelligence models by comparing cephalometric and panoramic radiographs. While most deep learning studies for dental age estimation have used panoramic radiographs, no study has directly compared cephalometric and panoramic modalities using the same deep learning architectures on the same patient cohort. This gap limits clinicians' ability to select the most informative imaging modality for AI-assisted dental age assessment. A head-to-head comparison under controlled experimental conditions is therefore scientifically necessary to clarify the relative utility of each modality and to provide important information for dental age prediction and clinical decision-making (Kim et al., 2021).

Research Objectives

This study aims to develop and evaluate artificial intelligence models that apply deep learning methods and transfer learning techniques to both cephalometric and panoramic radiographs for dental age estimation, with the following objectives:

1. To develop artificial intelligence models using deep learning models with transfer learning to predict dental age from cephalometric and panoramic radiographs.
2. To compare the performance of ResNet101 and EfficientNetB0 across both radiographic modalities and recommend appropriate model–modality combinations for dental age prediction.
3. To evaluate whether cephalometric radiographs can achieve prediction accuracy comparable to panoramic radiographs for dental age estimation.

Research Hypotheses

H1: Transfer-learning-based deep learning models can predict dental age from cephalometric radiographs with a Mean Absolute Error (MAE) of less than 2.5 years (using 2.5 years as a benchmark from panoramic-based studies by Kim et al., 2021 and Aljameel et al., 2023).

H2: Panoramic radiographs yield lower MAE than cephalometric radiographs for dental age estimation using the same model architecture and training conditions.

H3: ResNet101 and EfficientNetB0 demonstrate broadly comparable performance in terms of MAE across both radiographic modalities.

Literature Review

Dental age estimation has evolved significantly from traditional manual approaches to AI-driven methods. Conventional techniques such as the Demirjian and Cameriere methods have been widely used but are limited by their dependence on expert judgment, subjectivity, and time-intensive manual measurement (Brkic et al., 2022; Shen et al., 2021). The introduction of machine learning, and subsequently deep learning, has substantially improved accuracy and efficiency. Kim et al. (2021) demonstrated that automatic age estimation from panoramic radiographs using deep neural networks achieves competitive accuracy with reduced reliance on manual landmark identification. Mohamed et al. (2023) provided a comprehensive survey confirming that deep learning approaches consistently outperform traditional methods for dental age estimation. Collectively, these studies suggest that transfer-learning-based CNN architectures can achieve clinically meaningful error ranges, but the strongest evidence remains concentrated in panoramic imaging rather than cephalometric imaging.

1. Deep Learning Architectures Applied to Dental Radiographs

Several Convolutional Neural Network (CNN) architectures have been applied to dental age estimation with promising results. Aljameel et al. (2023) compared ResNet50, VGG16, Xception, and DenseNet121 for estimating canine age from panoramic radiographs, achieving a best MAE of 1.56 years with ResNet50, with further improvement when training data were restricted to narrower age ranges. Sributsayakarn et al. (2025) developed a multi-task EfficientNetB0 model for simultaneous age estimation and sex classification from Thai children's panoramic radiographs, achieving an MAE below 2–3 years. Oztekin et al. (2023) used neural networks with CBCT data, achieving an MAE of 4.12 years, substantially outperforming traditional regression models (MAE 8.17 years). Balel et al. (2025) demonstrated that deep learning-based systems using the Demirjian method on panoramic radiographs can further enhance estimation accuracy. These studies collectively show that transfer learning with architectures such as ResNet and EfficientNet is effective for dental age estimation, particularly with panoramic images.

However, nearly all existing deep learning studies for dental age estimation have focused on panoramic radiographs. Cephalometric radiographs, which provide lateral views enabling precise analysis of skeletal and dental relationships, remain underexplored for this purpose. Previous cephalometric studies have primarily addressed skeletal age, bone age, or anatomical landmark detection (Khanagar et al., 2021; Junaid et al., 2022), rather than dental age estimation per se. Singh and Raza (2022) noted that despite advances in deep learning for dental and maxillofacial image analysis, cephalometric applications remain limited, with small or imbalanced datasets and incomplete reporting of quantitative error metrics. Overall, prior studies indicate that deep learning models such as ResNet and EfficientNet can achieve promising performance for dental age estimation, particularly when panoramic radiographs are used. In contrast, cephalometric evidence remains limited and has focused more on skeletal or landmark-related tasks than on direct dental age prediction.

2. The Cross-Modality Gap

To date, no study has directly compared cephalometric and panoramic radiographs for dental age estimation using identical deep learning architectures trained on data from the same patient population and institution. This gap is significant because understanding the relative predictive value of each imaging modality would help clinicians and forensic practitioners select the most appropriate radiographic approach for AI-assisted dental age assessment. Given that cephalometric radiographs are routinely obtained in orthodontic practice and contain structural information about dental and skeletal development, investigating their utility for dental age estimation—and benchmarking their performance against the better-studied panoramic modality—addresses a clear and practical research need. The present study, therefore, fills this gap by conducting a controlled, head-to-head comparison under standardized experimental conditions. Therefore, a controlled comparison using the same architectures, dataset source, and training conditions is necessary to clarify whether cephalometric radiographs can provide performance comparable to panoramic radiographs in dental age estimation.

Research Methodology

Experimental research was conducted to develop and evaluate the performance of artificial intelligence (deep learning; AI) models for age prediction from dental radiographs. Two radiographic modalities were used: cephalometric X-rays and panoramic X-rays. The research process comprised (1) data preparation and cleaning, (2) data annotation/labeling, (3) model development and training, (4) evaluation, and (5) comparison of experimental results under the specified conditions: Classification—prediction of age classes/ranges; and Regression—prediction of age as a continuous value (age in years)

Figure 1: Overall Pipeline
Source: Author's Calculation

1. Methodology

1.1 Data Collection and Management

X-ray films were obtained from the hospital archive and categorized into two file types: cephalometric dental radiographs (Cephalometric X-ray) and panoramic radiographs (Panoramic X-ray). Images were required to be clear and to fully capture the dental and jaw structures to support the development of a model for predicting DA. The number of images was reported according to sex and chronological age. In addition, the dataset included clinical attributes such as age, sex, and tooth stage. This study was approved by the Kasetsart University Research Ethics Committee (Certificate of Exemption No. COE68/035), conducted in compliance with the Declaration of Helsinki, the Belmont Report, CIOMS Guidelines, and ICH-GCP (45 CFR 46.101(b)). All radiographic images were fully anonymized prior to use, with de-identified file codes assigned to protect patient privacy.

1.2 Dataset and Preprocessing

Table 1: Overall Dataset

No.	Image Type	Quantity	File Name (ID Name)	Gender	Age (year)	Source
1	Cephalometric Radiography	453	ID001DX_YX_M	Male	11-45 years	One Hospital Place
			ID001DX_YX_F	Female		
2	Panoramic Radiography	471	ID001PX_YX_M	Male		
			ID001PX_YX_F	Female		

For panoramic images, anonymized file codes were assigned to enable systematic data tracking while preserving patient privacy. The prepared images were uploaded to Roboflow and organized into two datasets: (1) cephalometric and (2) panoramic.

The data preparation steps were defined as follows:

(1) Resizing images to 640 x 640.

(2) Configuring data augmentation and, after verifying the settings, clicking Generate to create a new dataset version. Note: YOLOv11 and COCO format were used only in that preprocessing stage and during the data preparation and annotation validation phase within Roboflow's pipeline to verify image quality and annotation accuracy. YOLO was not used as a model for dental age prediction in this study; the age prediction models were exclusively ResNet101 and EfficientNetB0.

(3) Splitting the dataset into 70% training, 15% validation, and 15% test sets. The split was performed at the patient level, ensuring that images from the same individual appeared in only one subset to prevent data leakage. Although images were first organized at 640×640 pixels within the Roboflow preprocessing pipeline, they were resized to 224×224 pixels for model input during training in Google Colab.

1.3 Models and Training

The researchers prepared to export the dataset from Roboflow to Google Drive in order to run the models on Google Colab with GPU support. The dataset was downloaded from the Roboflow project page by selecting the COCO format and downloading it as a ZIP file. It was then uploaded to MyDrive and extracted into the Train/, Valid/, and Test/ directory structure to enable training in Colab. The candidate models were defined as CNN-based models for classification and regression, specifically ResNet101 and EfficientNetB0. Both models used ImageNet pre-trained weights with frozen base layers as the transfer learning strategy. The classification head used Categorical Cross-Entropy as the loss function, while the regression head used Mean Squared Error (MSE) loss. Both models were trained with the Adam optimizer (learning rate = 0.0001) and a batch size of 16. For the classification task, the continuous age range (11–45 years) was divided into 7 age classes using 5-year bins (11–15, 16–20, 21–25, 26–30, 31–35, 36–40, 41–45). The model head was designed separately as a classification head and a regression head. The experimental design specified the main factors as image type (cephalometric vs. panoramic) and 2 models. A repeated-run protocol was used by executing the same steps across all conditions (image type × augmentation × epoch × model × task) to enable systematic comparison of results.

Table 2: Models and Factors Used for Model Training and Testing

No.	Image Type	Model	epoch		Augmentation	
1	Cephalometric X-ray	ResNet101	10	30	No	Add
2	Cephalometric X-ray	EfficientNetB0	10	30	No	Add
3	Panoramic X-ray	ResNet101	10	30	No	Add
4	Panoramic X-ray	EfficientNetB0	10	30	No	Add

1.4 Evaluation and Model Selection

Model selection criteria were defined according to the task. For classification, the Weighted F1-score or the Macro F1-score was used as the primary metric, with accuracy as a supporting metric. For regression, models were selected based on the lowest MAE/RMSE and consistent performance in testing.

1) Accuracy
2) Precision, Recall, F1-score
3) MAE and RMSE

Results

The overall findings summarize an experimental study on age prediction using two types of X-ray images: cephalometric and panoramic X-rays, with ResNet101 and EfficientNetB0 under 16 conditions combining epochs (10/30) and augmentation (No/Add). Panoramic X-rays clearly yielded better performance than cephalometric X-rays, with an average MAE of ~3.15 years and an average RMSE of ~4.26 years. In contrast, cephalometric X-rays achieved an average MAE of ~3.87 years and an average RMSE of ~5.48 years. For classification evaluation, test accuracy was constant across all conditions, at 0.4038 for cephalometric images and 0.5000 for panoramic images, for both 10 and 30 training epochs. This constant classification accuracy warrants careful interpretation. The age range of 11–45 years was divided into 7 classes using 5-year bins, but the distribution of samples across these bins was uneven, with certain age groups having substantially more samples than others. Combined with

the relatively small test set (~68 cephalometric and ~71 panoramic images), the models likely converged to a majority-class prediction strategy, producing identical accuracy regardless of hyperparameter variation. This pattern indicates that the classification formulation did not meaningfully differentiate across experimental conditions, and the regression-based evaluation (MAE, RMSE) provides a more reliable basis for model comparison in this study.

Table 3: Summary of Test Performance (Accuracy, Loss, Precision, Recall, F1-Score, MAE, RMSE) for Cephalometric and Panoramic X-ray Models

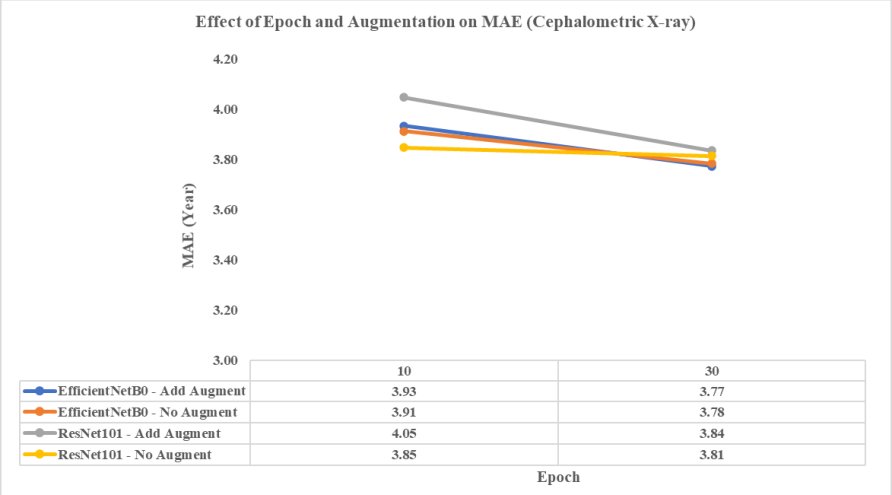
Trial	Image Type (X-ray Film)	Model	Epoch	Augmentation	Test Accuracy	Test Loss	Precision	Recall	F1-Score	MAE (years)	RMSE (years)
1	Cephalometric	ResNet101	10	No Augment	0.4038	1.4430	0.163	0.404	0.232	3.85	5.49
2	Cephalometric	EfficientNetB0	10	No Augment	0.4038	1.4310	0.163	0.404	0.232	3.91	5.47
3	Panoramic	ResNet101	10	No Augment	0.5000	1.2770	0.250	0.500	0.333	3.04	4.27
4	Panoramic	EfficientNetB0	10	No Augment	0.5000	1.2840	0.250	0.500	0.333	3.21	4.28
5	Cephalometric	ResNet101	30	No Augment	0.4038	1.4650	0.163	0.404	0.232	3.81	5.48
6	Cephalometric	EfficientNetB0	30	No Augment	0.4038	1.4590	0.163	0.404	0.232	3.78	5.51
7	Panoramic	ResNet101	30	No Augment	0.5000	1.2440	0.250	0.500	0.333	3.03	4.25
8	Panoramic	EfficientNetB0	30	No Augment	0.5000	1.2490	0.250	0.500	0.333	3.12	4.27
9	Cephalometric	ResNet101	10	Add Augment	0.4038	1.4440	0.163	0.404	0.232	4.05	5.47
10	Cephalometric	EfficientNetB0	10	Add Augment	0.4038	1.4420	0.163	0.404	0.232	3.93	5.47
11	Panoramic	ResNet101	10	Add Augment	0.5000	1.2590	0.250	0.500	0.333	3.25	4.18
12	Panoramic	EfficientNetB0	10	Add Augment	0.5000	1.2840	0.250	0.500	0.333	3.32	4.29
13	Cephalometric	ResNet101	30	Add Augment	0.4038	1.4386	0.163	0.404	0.232	3.84	5.48
14	Cephalometric	EfficientNetB0	30	Add Augment	0.4038	1.4556	0.163	0.404	0.232	3.77	5.50
15	Panoramic	ResNet101	30	Add Augment	0.5000	1.2400	0.250	0.500	0.333	3.09	4.26
16	Panoramic	EfficientNetB0	30	Add Augment	0.5000	1.2480	0.250	0.500	0.333	3.11	4.27

Source: (Author's Calculation)

1. Overall Performance

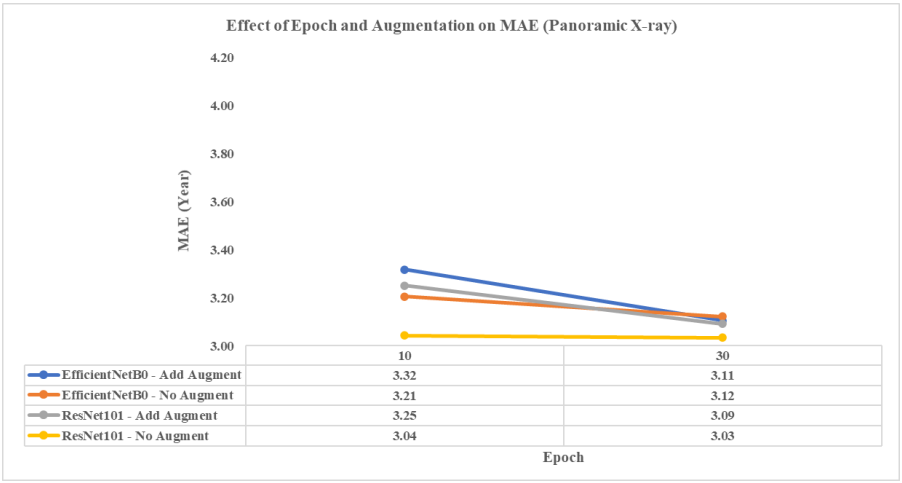
Overall performance indicates that panoramic X-rays outperformed cephalometric X-rays, yielding lower MAE and RMSE across all conditions. ResNet101 and EfficientNetB0 showed comparable overall results. Increasing the number of epochs from 10 to 30 helped improve regression performance by reducing MAE. Data augmentation produced mixed outcomes, with its effect depending on the image modality, model, and training settings from the graph. It presents the effects of the number of training epochs (Epoch 10 vs 30) and data augmentation on the MAE for dental age prediction from Panoramic X-ray and cephalometric X-ray images by using ResNet101 and EfficientNetB0.

1.1 Effect of Epoch and Augmentation on MAE Cephalometric and Panoramic X-ray



	10	30
EfficientNetB0 - Add Augment	3.93	3.77
EfficientNetB0 - No Augment	3.91	3.78
ResNet101 - Add Augment	4.05	3.84
ResNet101 - No Augment	3.85	3.81

Figure 2: Effect of Epoch and Augmentation on MAE (Cephalometric X-ray)
Source: (Author's Calculation)



	10	30
EfficientNetB0 - Add Augment	3.32	3.11
EfficientNetB0 - No Augment	3.21	3.12
ResNet101 - Add Augment	3.25	3.09
ResNet101 - No Augment	3.04	3.03

Figure 3: Effect of Epoch and Augmentation on MAE (Panoramic X-ray)
Source: (Author's Calculation)

The results from the graphs in Figures 2 and 3 indicate that increasing the number of training epochs from 10 to 30 led to a clear reduction in MAE for panoramic X-rays. MAE was approximately 3.03–3.32 years, and decreased notably in the augmented groups from 3.3 to 3.11 years and 3.25 to 3.09 years. For cephalometric X-rays, MAE was higher by approximately 3.77-4.05 years, but still decreased with more epochs, particularly for ResNet101 with augmentation from 4.05 to 3.84 years and EfficientNetB0 with augmentation from 3.93 to 3.77 years. Overall, increasing the number of epochs was the factor most strongly associated with a reduction in MAE. In contrast, augmentation served as an additional factor that may improve performance under specific training conditions and can be used as a supplementary approach to enhance model performance when appropriate.

1.2 Comparison of MAE (year) by Model of Cephalometric and Panoramic X-ray

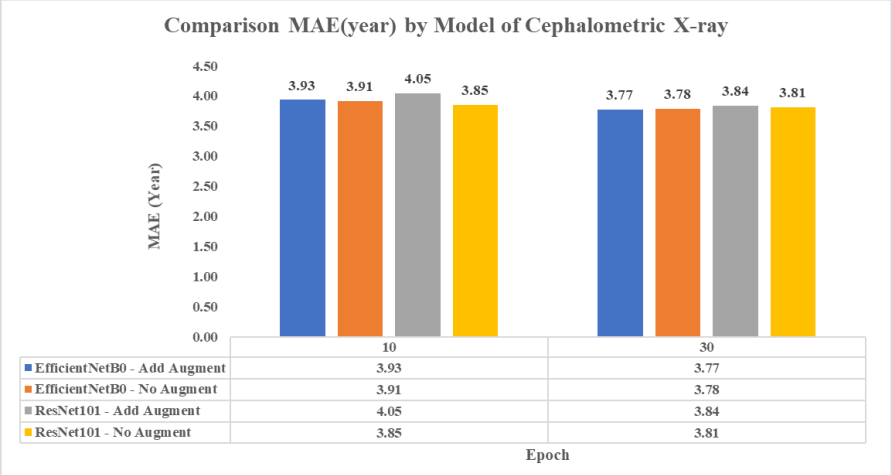


Figure 4: Comparison MAE (year) by Model of Cephalometric X-ray
Source: (Author's Calculation)

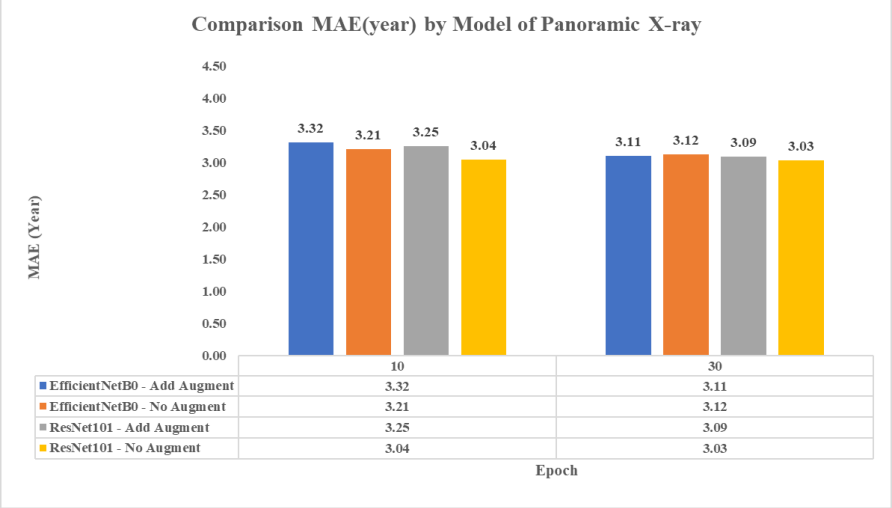


Figure 5: Comparison MAE (year) by Model of Panoramic X-ray
Source: (Author's Calculation)

The experimental results from Figures 4 and 5 show that increasing the number of epochs from 10 to 30 reduced MAE for cephalometric images under all conditions. For EfficientNetB0 with augmentation, MAE decreased from 3.93 to 3.77 years, while for ResNet101 with augmentation, MAE decreased from 4.05 to 3.84 years. This indicates that the models can learn and improve prediction accuracy with longer training, consistent with the objective of developing models using transfer learning. The graphs also provide a basis for comparing models to select an appropriate one. Overall, ResNet101 and EfficientNetB0 achieved comparable performance, but the most suitable conditions varied with the complexity of the X-ray images.

1.3 Effect of Data Augmentation on Age Prediction Error (ΔMAE) in Cephalometric and Panoramic X-ray

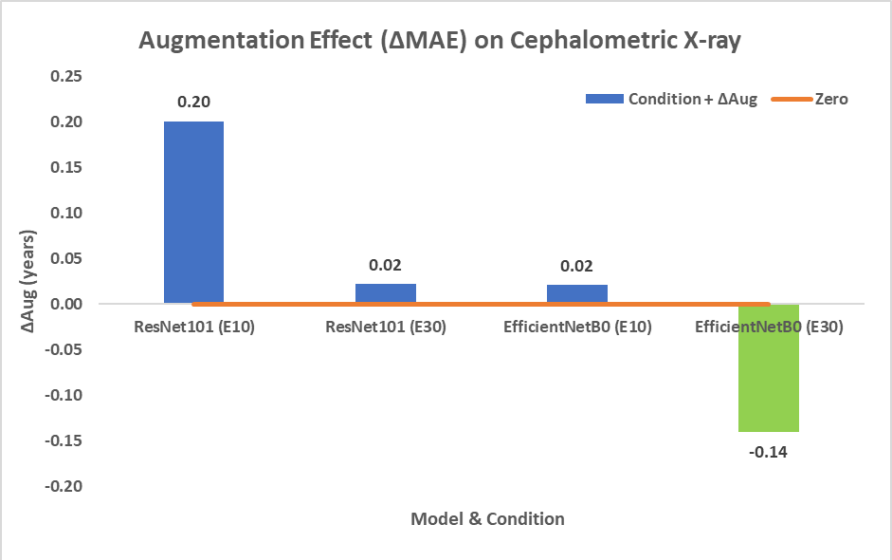


Figure 6: Augmentation Effect (ΔMAE) on Cephalometric X-ray
Source: (Author's Calculation)

Results from Figure 6 for cephalometric X-rays, ResNet101 at 10 epochs showed the greatest deterioration when augmentation was applied, with an error of ΔMAE = +0.20. ResNet101 at 30 epochs and EfficientNetB0 at 10 epochs showed a slight deterioration (ΔMAE = +0.02). In contrast, EfficientNetB0 at 30 epochs showed the most pronounced improvement, with ΔMAE = □0.14. It can therefore be concluded that augmentation is most effective when applied to EfficientNetB0 at 30 epochs.

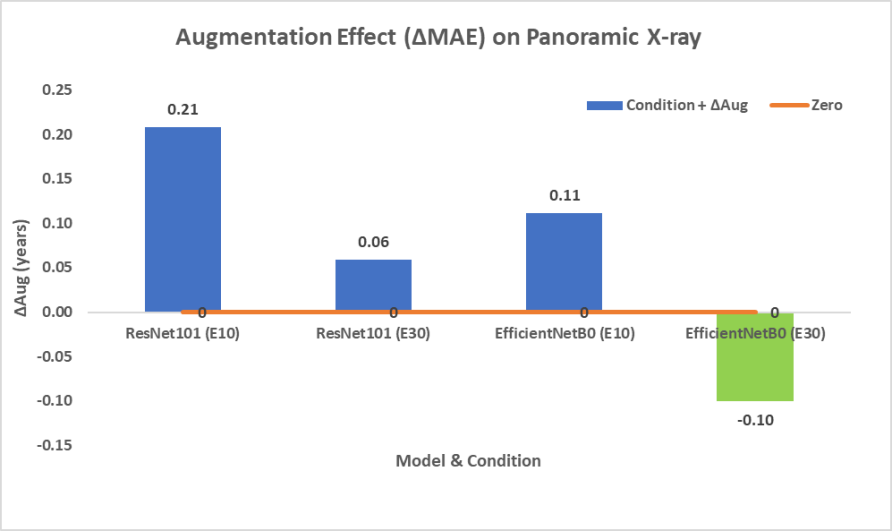


Figure 7: Augmentation Effect (ΔMAE) on Panoramic X-ray
Source: (Author's Calculation)

Experimental results from Figure 7, ResNet101 performed worse when augmentation was applied. At 10 epochs, the error increased (ΔMAE = +0.21), and at 30 epochs, the error was ΔMAE = +0.06. EfficientNetB0 at 10 epochs showed only a slight deterioration (ΔMAE = +0.11), whereas EfficientNetB0 at 30 epochs showed an improvement in error (ΔMAE = □0.10).

4. Effect of Increasing the Number of Training Epochs on Age Prediction Error (ΔMAE) in Cephalometric and Panoramic X-ray

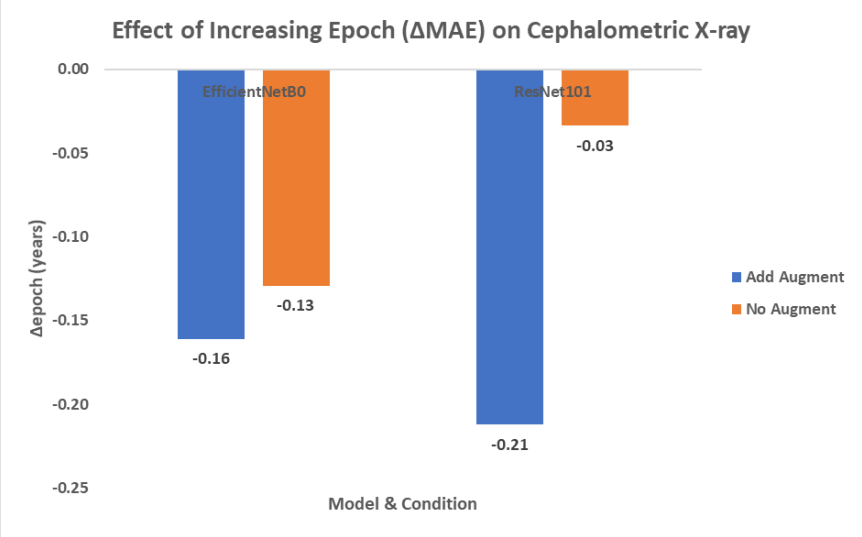


Figure 8: Effect of Increasing Epoch (ΔMAE) on Cephalometric X-ray
Source: (Author's Calculation)

Experimental results from Figure 8 showed that increasing the number of epochs for EfficientNetB0 reduced MAE under both conditions: Add Augment (□0.16) and No Augment (□0.13). For ResNet101, increasing epochs reduced MAE more clearly when augmentation was applied, with the largest reduction under Add Augment (-0.21), whereas without augmentation, the reduction was minimal (-0.03). Concluded that, for cephalometric X-rays, increasing the number of epochs to 30 helps reduce MAE.

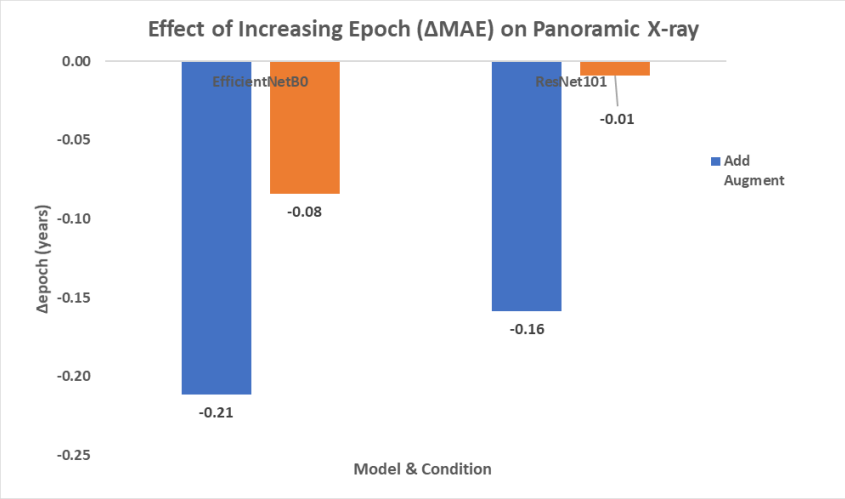


Figure 9: Effect of Increasing Epoch (ΔMAE) on Panoramic X-ray
Source: (Author's Calculation)

Experimental results from Figure 9 for EfficientNetB0 show that increasing the number of epochs reduced MAE, with the largest reduction under the Add Augment condition (□0.21), whereas the reduction under No Augmentation was only □0.08. For ResNet101, increasing epochs led to only a

slight reduction in MAE, with 0.16 under Add Augmentation and a negligible change under No Augmentation (0.01). Therefore, for panoramic X-rays, the effect of increasing epochs is more evident with the EfficientNetB0 model.

Discussion

The selection and use of appropriate artificial intelligence models for dental age prediction were investigated using two types of X-ray images: panoramic and cephalometric radiographs. ResNet101 and EfficientNetB0 were evaluated under 16 experimental conditions. The results confirmed that panoramic radiographs consistently produced lower prediction errors than cephalometric radiographs, with the best panoramic MAE of 3.03 years (ResNet101, 30 epochs, no augmentation) compared to the best cephalometric MAE of 3.77 years (EfficientNetB0, 30 epochs, with augmentation). This difference is likely attributable to the inherent information content of each modality: panoramic radiographs capture all teeth and both dental arches in a single frontal image, providing richer morphological information directly relevant to age-related dental changes, whereas cephalometric radiographs provide a lateral view in which dental structures overlap and are less individually distinguishable, making age-relevant features harder for the model to extract.

Hypothesis H1, which proposed that cephalometric models could achieve MAE below 2.5 years, was not supported. The lowest cephalometric MAE (3.77 years) exceeded this threshold by 1.27 years. This finding is itself a meaningful contribution, as it establishes a quantitative benchmark for cephalometric-based dental age estimation and highlights the greater difficulty of this task compared with panoramic-based estimation. The 2.5-year threshold was derived from panoramic-based studies (Kim et al., 2021; Aljameel et al., 2023), and the present results confirm that this threshold may not be directly transferable to cephalometric images without further methodological refinement.

Hypothesis H2 was supported: panoramic radiographs yielded lower MAE than cephalometric radiographs across all experimental conditions, confirming the cross-modality performance difference.

Hypothesis H3 was also supported: ResNet101 and EfficientNetB0 demonstrated comparable overall performance, although the optimal model–modality pairing differed. For cephalometric images, EfficientNetB0 performed best, while for panoramic images, ResNet101 achieved the lowest MAE. This suggests that model selection should be modality-dependent.

The constant classification accuracy (0.4038 for cephalometric; 0.5000 for panoramic) across all 16 conditions indicates that the classification task, as formulated with 7 age-class bins, did not produce meaningful differentiation. This is further confirmed by the constant precision (0.163/0.250), recall (0.404/0.500), and F1-score (0.232/0.333) values for cephalometric and panoramic images, respectively, across all conditions. This is likely due to class imbalance within the 5-year bins combined with the small test set size. Future work should consider alternative approaches such as broader age groupings, ordinal regression, or focusing exclusively on regression-based evaluation.

Compared with prior studies, the present results fall within a plausible range given the study design. Aljameel et al. (2023) achieved an MAE of 1.56 years using ResNet50 on panoramic radiographs, but with a narrower age range (6–12 years). Sributsayakarn et al. (2025) reported MAE below 2–3 years using EfficientNetB0 on Thai panoramic radiographs for ages 7–23 years. The wider age range in the present study (11–45 years) and the smaller dataset size (924 images) likely contribute to the higher MAE values. Oztekin et al. (2023) reported an MAE of 4.12 years using neural networks on CBCT data for a

14–60-year age range, which is comparable to the cephalometric results obtained here. These comparisons suggest that both age range breadth and imaging modality substantially influence prediction accuracy.

Increasing the number of training epochs from 10 to 30 consistently reduced MAE across most conditions, indicating that the models benefited from additional training iterations within the transfer learning framework. Data augmentation produced mixed results, improving performance in some conditions (particularly EfficientNetB0 at 30 epochs) while slightly degrading it in others (particularly ResNet101 at 10 epochs). This suggests that augmentation is most beneficial when combined with sufficient training duration, and its effect is model- and modality-dependent. These findings address Objective 2 by providing empirical evidence for model selection: EfficientNetB0 with augmentation is recommended for cephalometric images, while ResNet101 without augmentation is recommended for panoramic images.

Conclusion

This study compared ResNet101 and EfficientNetB0 for dental age estimation using cephalometric and panoramic radiographs under 16 experimental conditions. The key findings are as follows. First, panoramic radiographs consistently outperformed cephalometric radiographs, achieving a best MAE of 3.03 years (ResNet101, 30 epochs, no augmentation) compared to 3.77 years for cephalometric images (EfficientNetB0, 30 epochs, with augmentation), supporting Hypothesis H2. Second, Hypothesis H1 was not supported: the proposed MAE threshold of 2.5 years for cephalometric images was not achieved under any condition, indicating that cephalometric-based dental age estimation remains less accurate than panoramic-based estimation with the current architectures and datasets. Third, ResNet101 and EfficientNetB0 showed comparable overall performance (supporting H3), but the optimal model differed by modality. Fourth, the classification task (7 age classes in 5-year bins) did not produce meaningful differentiation across conditions, with constant test accuracy of 0.4038 (cephalometric) and 0.5000 (panoramic), suggesting that the regression framework is the more informative evaluation approach for this task. These findings provide practical guidance for selecting appropriate imaging modality and model architecture for AI-assisted dental age estimation.

Limitations and Future Work

This study has several limitations that should be acknowledged. First, the dataset was collected from a single healthcare facility (924 images), which may limit generalizability to other populations or imaging equipment. Second, the classification task did not yield meaningful variation across conditions, suggesting that the 7-class binning scheme with uneven class distributions was suboptimal. Third, each experimental condition was run once; repeated runs with different random seeds would be needed to assess the robustness and variability of the reported results. Fourth, precision, recall, and F1-score have been added to Table 3, confirming the constant classification performance pattern. Although confusion matrices were examined, they were not emphasized in the main text because the models showed highly concentrated majority-class predictions across conditions, providing limited additional insight beyond the constant classification metrics, as the constant classification accuracy across all conditions indicates that models converged to majority-class prediction, rendering per-class confusion analysis uninformative under the current experimental design. Fifth, the study did not include statistical tests (e.g., confidence intervals or paired comparisons) to assess whether the observed MAE differences between conditions are statistically significant.

Future work should address these limitations by: (a) expanding the dataset to multiple institutions and populations to improve generalizability; (b) exploring narrower age-range stratification or ordinal regression to improve classification performance; (c) investigating alternative architectures such as Vision Transformers (ViT) or ensemble methods combining both modalities; (d) conducting formal statistical comparisons between experimental conditions; and (e) benchmarking deep learning results against traditional dental age estimation methods (e.g., Demirjian, Cameriere) on the same dataset. These steps would strengthen the evidence base before clinical or forensic application can be recommended.

References

Aljameel, S. S., Althumairy, L., Albassam, B., Alsheikh, G., Albluwi, L., Althukair, R., Alhareky, M., Alamri, A., Alabdan, A., & Shahin, S. Y. (2023). Predictive Artificial Intelligence Model for Detecting Dental Age Using Panoramic Radiograph Images. *Big Data and Cognitive Computing*, *7*(1). https://doi.org/10.3390/bdcc7010008

Balel, Y., Sagtas, K., & Bulbul, H. N. (2025). Development and evaluation of a deep learning-based system for dental age estimation using the demirjian method on panoramic radiographs. *BMC Oral Health*, *25*(1), 1172. https://doi.org/10.1186/s12903-025-06420-5

Brahmi, W., Jdey, I., & Drira, F. (2024). Exploring the role of Convolutional Neural Networks (CNN) in dental radiography segmentation: A comprehensive Systematic Literature Review. *Engineering Applications of Artificial Intelligence*, *133*, 108510. https://doi.org/10.1016/j.engappai.2024.108510

Brkic, H., Galic, I., Vodanovic, M., Dumancic, J., Mehdi, F., & Anic Milosevic, S. (2022). The Cameriere, Haavikko, Demirjian, and Willems methods for the assessment of dental age in Croatian children. *Int J Legal Med*, *136*(6), 1685-1696. https://doi.org/10.1007/s00414-022-02891-1

Franco, A., Cornacchia, A. P., Moreira, D., Miamoto, P., Bueno, J., Murray, J., Heng, D., Manica, S., Porto, L., & Abade, A. (2024). Radiographic morphology of canines tested for sexual dimorphism via convolutional-neural-network-based artificial intelligence. *Morphologie*, *108*(362), 100772. https://doi.org/10.1016/j.morpho.2024.100772

Junaid, N., Khan, N., Ahmed, N., Abbasi, M. S., Das, G., Maqsood, A., Ahmed, A. R., Marya, A., Alam, M. K., & Heboyan, A. (2022). Development, Application, and Performance of Artificial Intelligence in Cephalometric Landmark Identification and Diagnosis: A Systematic Review. *Healthcare*, *10*(12). https://doi.org/10.3390/healthcare10122454

Juneja, M., Garg, P., Kaur, R., Manocha, P., Prateek, Batra, S., Singh, P., Singh, S., & Jindal, P. (2021). A review on cephalometric landmark detection techniques. *Biomedical Signal Processing and Control*, *66*. https://doi.org/10.1016/j.bspc.2021.102486

Khanagar, S. B., Vishwanathaiah, S., Naik, S., A, A. A.-K., Devang Divakar, D., Sarode, S. C., Bhandi, S., & Patil, S. (2021). Application and performance of artificial intelligence technology in forensic odontology - A systematic review. *Legal Medicine*, *48*, 101826. https://doi.org/10.1016/j.legalmed.2020.101826

Kim, E. G., Oh, I. S., So, J. E., Kang, J., Le, V. N. T., Tak, M. K., & Lee, D. W. (2021). Estimating Cervical Vertebral Maturation with a Lateral Cephalogram Using the Convolutional Neural Network. *Journal of Clinical Medicine*, *10*(22), 5400. https://doi.org/10.3390/jcm10225400

Koch, R. M., Mentzel, H. J., & Heinrich, A. (2025). Deep learning for forensic age estimation using orthopantomograms in children, adolescents, and young adults. *Eur Radiol*, *35*(7), 4191-4202. https://doi.org/10.1007/s00330-025-11373-y

Lee, T., Shin, W., Lee, J. H., Lee, S., Yeom, H. G., & Yun, J. P. (2024). Resolving the non-uniformity in the feature space of age estimation: A deep learning model based on feature clusters of panoramic images. *Comput Med Imaging Graph*, *112*, 102329. https://doi.org/10.1016/j.compmedimag.2024. 102329

Mohamed, E. G., Redondo, R. P. D., Koura, A., El-Mofty, M. S., & Kayed, M. (2023). Dental Age Estimation Using Deep Learning: A Comparative Survey. *Computation*, *11*(2). https://doi.org/10. 3390/computation11020018

Mohammad, N., Ahmad, R., Gaus, M. H. A., Kurniawan, A., & Yusof, M. (2024). Accuracy of automated forensic dental age estimation lab (F-DentEst Lab) on large Malaysian dataset. *Forensic Science International*, *361*, 112150. https://doi.org/10.1016/j.forsciint.2024.112150

Oztekin, F., Katar, O., Sadak, F., Yildirim, M., Cakar, H., Aydogan, M., Ozpolat, Z., Talo Yildirim, T., Yildirim, O., Faust, O., & Acharya, U. R. (2023). An Explainable Deep Learning Model to Prediction Dental Caries Using Panoramic Radiograph Images. *Diagnostics*, *13*(2). https://doi.org/10.3390/diagnostics13020226

Shen, S., Liu, Z., Wang, J., Fan, L., Ji, F., & Tao, J. (2021). Machine learning assisted Cameriere method for dental age estimation. *BMC Oral Health*, *21*(1), 641. https://doi.org/10.1186/s12903-021-01996-0

Singh, N. K., & Raza, K. (2022). Progress in deep learning-based dental and maxillofacial image analysis: A systematic review. *Expert Systems with Applications*, *199*. https://doi.org/10.1016/j.eswa.2022.116968

Sributsayakarn, N., Intharah, T., Hirunchavarod, N., Pornprasertsuk-Damrongsri, S., & Jirarattanasopha, V. (2025). Automated Age and Sex Estimation From Dental Panoramic Radiographs. *International Dental Journal*, *75*(6), 103967. https://doi.org/10.1016/j.identj.2025.103967

EXPLORING AI AWARENESS: A COMPREHENSIVE REVIEW AND FUTURE OUTLOOK

Fangfang Zhou[1*], Zuju Yao[2], Rui Mao[3], and Lei Zhou[4]

[1]International Chinese College, Panyapiwat Institute of Management, Thailand
[2,3,4]School of Business Administration, Guizhou University of Finance and Economics, China
[*]**Corresponding Author**: 740442198@qq.com

Abstract

A single robot can replace 70 full-time employees. Moreover, AI and robots will replace 47% of jobs in the future. Artificial intelligence (AI) awareness refers to employees' subjective perception of uncertainty regarding their future career development and work environment, stemming from the recognition that AI technologies may potentially replace their current job positions. This study systematically reviews, screens, and analyzes the literature on AI awareness, synthesizing its definitions and measurement approaches. It also identifies the main consequences of AI awareness across three key dimensions: employees' psychological states, behaviors, and job performance. Based on these findings, the study suggests that future research should focus on uncovering the formation mechanisms of AI awareness, exploring its antecedents, broadening the range of its outcome variables, and enriching the theoretical foundations underpinning its impact pathways. These insights aim to provide scholars with a comprehensive understanding of research progress in the field of AI awareness, while also assisting managers in better understanding and effectively managing employees' perceptions of AI.

Keywords: AI Awareness, Robot, Job Insecurity, Turnover Intention

Introduction

Artificial intelligence (AI) refers to a technology that emulates human intelligence through computer systems, encompassing capabilities such as learning, reasoning, perception, and language processing. AI has found extensive applications across diverse industries, including medical imaging analysis in healthcare (Karalis, 2024), intelligent tutoring systems in education (Rawas, 2024), and performance enhancement in aerospace (Hassan et al., 2024). Furthermore, AI plays a crucial role in revenue management, marketing, engineering management, and human resources (Krabokoukis, 2023; Shen, 2022). To summarize the above, we have compiled the usage of AI across various industries, as shown in Table 1. Compared to other sectors, the hospitality industry has witnessed particularly widespread adoption of AI, which is rapidly transforming its landscape (He et al., 2024). Frontline hotel employees typically perform repetitive and routine tasks with low entry barriers, making their roles highly susceptible to replacement by robots specialized in executing standardized operations (Arias-Pérez & Vélez-Jaramillo, 2022a; Teng et al., 2024). Additionally, customers increasingly expect advanced technological solutions and personalized experiences, prompting a growing number of hotels to integrate sophisticated robotics and AI throughout multiple stages of consumer service (Shum et al., 2024). During the customer inquiry phase, AI-powered chatbots can promptly address customer queries regarding hotel facilities, services, and booking policies while facilitating reservation procedures (Koo et al., 2021). In customer service, AI analyzes customer preferences prior to arrival to generate personalized recommendations, such as suggesting suitable room types, dining options, and activities based on historical stay records, while AI-driven revenue management systems dynamically adjust room rates by analyzing real-time market demand, competitor pricing, and seasonal factors (Arias-Pérez & Vélez-Jaramillo, 2022b; Parvez et al., 2022). During the stay, AI

assists the front desk or mobile applications by providing real-time responses to geographical and facility-related inquiries, thereby reducing wait times. It further enables self-service check-in and check-out processes to minimize queues, while offering multilingual support. Concurrently, the system analyzes customer consumption patterns and behavioral trajectories within the premises to deliver personalized services and recommendations (He et al., 2024). During the post-departure phase of the customer, AI evaluates operational data including occupancy rates, pricing, customer satisfaction, and marketing effectiveness to support managerial decision-making (Parvez et al., 2022).

Currently, AI and robotics technologies have been extensively implemented in hotels worldwide. For instance, properties such as Wynn Las Vegas and Aloft Hotels in the United States have adopted virtual assistants to handle in-room customer requests (Li et al., 2019), while Japan's Henn-na Hotel utilizes robotic porters, cloakroom robots, and in-room personal assistants (Kong et al., 2021). As of 2020, Chinese hospitality groups had deployed AI technologies across more than 5,700 hotels (Lin & He, 2024). Robots demonstrate significantly faster operational speeds compared to human workers, with a single robot capable of replacing 70 full-time employees. The application of AI and robotic technologies not only meets customers' personalized demands, enhances satisfaction, and optimizes service delivery, but also offers advantages such as improving organizational efficiency and reducing labor costs (Yin et al., 2024). These benefits have accelerated the hotel industry's rapid shift toward intelligent development characterized by the replacement of human labor with AI. Relevant studies indicate that frontline hotel staff face up to 95% probability of job displacement by robots (Frey & Osborne, 2017), with global estimates suggesting 400-800 million positions may be at risk of automation by 2030 (Parvez et al., 2022). This profound transformation has led hotel employees to gradually develop AI awareness, referring to their subjective perception that artificial intelligence may impact their future career development (Teng et al., 2024).

In the process of modernizing organizational management in the hotel industry, AI awareness has become a critical topic in both practical exploration and theoretical research. However, current studies have largely focused on empirical investigations of AI awareness, while relatively few scholars have conducted comprehensive reviews and summaries of the literature in this field. It is therefore necessary to systematically and objectively examine the existing body of research on AI awareness in order to gain a holistic understanding of its academic development and emerging trends (Bakir et al., 2025).Based on this consideration, the present study undertakes several main tasks. First, through a series of procedures including database searching, initial screening, evaluation, and selection, thirty relevant articles were identified as the analytical sample. Second, the study analyzes the definition and measurement of AI awareness, aiming to clarify and summarize the conceptual connotations and measurement approaches used in the existing literature. This analysis not only helps assess the consistency and precision of the concept's application in academic studies but also supports practical efforts to identify and evaluate employees' awareness of AI in organizational settings.In addition, the study investigates the theoretical framework of AI awareness by exploring its antecedents and consequences. This approach contributes to uncovering the deeper theoretical mechanisms underlying the formation and outcomes of AI awareness and offers a way to overcome the limitations of existing review studies. Finally, based on a comprehensive analysis of related research, the study outlines future research directions in this field.The theoretical value and contributions of this study can be summarized in two main aspects. On the one hand, it provides important insights and guidance for advancing theoretical research on AI awareness and helps hotel managers develop a more accurate understanding of the nature of AI awareness, thereby preventing or mitigating its potential negative impacts. On the other hand, by proposing an integrated research framework and identifying future directions, this study encourages scholars to explore related issues from multiple perspectives, thereby enriching both theoretical understanding and practical application.

Table 1: The application of AI in various industries

Classification of Industry	Occupancy Ratio
Travel	45%
Financial Service	40%
Health Care	35%
Manufacturing Industry	30%
Hotel	27%
Retail Industry	25%
Transportation	20%
Education	15%
Energy	10%
Agriculture	5%
Media and Entertainment	5%

Content

1. The conceptualization of AI awareness

Currently, there is a relatively consistent understanding among scholars regarding the concept of AI awareness.It refers to employees' recognition that robots, algorithms, and other AI technologies may potentially replace their current jobs, which leads to a perception of uncertainty regarding the workplace in the digital era. (Brougham & Haar, 2018). Kong et al. (2021) explained that with the rapid advancement of AI technologies, their potential to outperform humans in various complex tasks and to replace human labor in many jobs has triggered employees' concerns about being replaced. These concerns contribute to a sense of uncertainty regarding their future career prospects.Teng et al. (2024) define AI awareness as employees' cognitive perception of AI's impact on their future work and career prospects. Bakir et al. (2025) complement this conceptualization by emphasizing employees' subjective assessment of the likelihood that AI and robotic technologies might displace their current job positions.

2. Measurement of AI awareness

Current research demonstrates relative consistency in the measurement of AI awareness, primarily employing the 4-item scale adapted by Brougham and Haar (2018) from Armstrong-Stassen's (2001) job insecurity instrument. Armstrong-Stassen (2001) defined job insecurity as consisting of two distinct dimensions: perceived threat of job loss and feelings of powerlessness. The former includes three items, while the latter is measured by one item. This approach emphasizes assessing AI awareness by simulating human experiences of job insecurity, which may involve evaluating an AI system's recognition of its own existence or value and its perceived uncertainty (Barnes & Hutson, 2024). Initially developed to measure STARA (smart technology, AI, robotics, and algorithms) awareness (Kang et al., 2023; Zhao et al., 2023; Yang et al., 2024), Brougham and Haar(2018)'s measurement scale has subsequently emerged as the predominant instrument for operationalizing AI awareness in empirical studies (Kong et al., 2021; Li et al., 2019).

3. Consequences of AI awareness

Recent studies have focused on the results of AI awareness, mainly concentrating on its impact on employees' personal states, their behaviors, and their performance. As shown in Figure 1. In addition, the research years and countries related to AI awareness have also been compiled, as shown in Table 2.

In terms of employees' personal states, scholars have focused on aspects such as turnover intention, work engagement, job burnout, health symptoms, capabilities, and workplace well-being.First, scholars have found that AI awareness influences turnover intention. For example,drawing on the stress-emotion model, conservation of resources theory, and social exchange theory, Doan and Nguyen

(2025) found that AI awareness positively influences turnover intention through the mediating role of interpersonal conflict. In this process, a competitive psychological climate and perceived organizational support act as moderating variables. Similarly, based on conservation of resources theory and social exchange theory, scholars have shown that AI awareness has a positive effect on turnover intention, with supervisor support moderating the relationship between AI awareness and turnover intention, while the moderating effect of a competitive psychological climate is not significant (Moin & Zhang, 2024). Li et al. (2019) also revealed that awareness of AI and robotics positively affects turnover intention, and that perceived organizational support and a competitive psychological climate moderate the relationship between AI awareness and turnover intention. Supporting this view, based on conservation of resources theory, Khaliq et al. (2022) found that AI and robotics awareness positively influence turnover intention. Mutual trust moderates the relationship between AI awareness and turnover intention, whereas the moderating effect of a competitive psychological climate is not significant. In addition, based on person-environment fit theory and social exchange theory, Kang et al. (2024) found that AI awareness positively influences turnover intention toward both other hotels and non-hotel industries through the mediating roles of job stress and job insecurity. Social capital was found to moderate only the effect of job insecurity on turnover intention toward non-hotel industries.Second, AI awareness has been shown to influence both work engagement and job burnout. Drawing on conservation of resources theory and the job demand-resources model, Liu and Cheng (2025) found that AI awareness positively affects work engagement through the mediating roles of recovery level and job insecurity. They further demonstrated that innovativeness as a job requirement moderates the relationships between AI awareness and both job insecurity and recovery level. Zheng and Zhang (2025) revealed that AI awareness exacerbates emotional exhaustion through a chain mediation involving job insecurity and work-family conflict.Integrating job crafting theory and social exchange theory, Lestari et al. (2024) found that both AI awareness and servant leadership positively influence work engagement. Specifically, servant leadership can enhance employee engagement indirectly through job crafting, although job crafting does not play a significant mediating role between AI awareness and work engagement. In contrast, based on conservation of resources theory, Kong et al. (2025) found that AI awareness negatively affects work engagement through the mediation of job insecurity, and that self-efficacy moderates this relationship. In addition, drawing on job stress theory, Khairy et al. (2024) found that AI awareness negatively influences green work engagement, with job stress mediating this relationship. Technological self-efficacy and trust in leadership respectively moderate the relationship between AI awareness and job stress, and the relationship between job stress and green work engagement. Furthermore, AI awareness has been found to increase job burnout, and this effect becomes more pronounced among employees with higher levels of organizational commitment (Kong et al., 2021).Third, AI awareness affects employees' health symptoms. Based on conservation of resources theory, Xu et al. (2023) found that AI awareness increases employee depressive symptoms by exacerbating emotional exhaustion. Simultaneously, high perceived organizational support can attenuate the effect of AI awareness on depressive symptoms via emotional exhaustion. Drawing on the transactional theory of stress and coping, other scholars revealed that AI awareness positively influences mental health symptoms, with role conflict acting as a mediator in this process. Furthermore, AI-oriented human resource analytics and human resource management system strength the effect of AI awareness on mental health symptoms through role conflict (Yan et al., 2025).Fourth, AI awareness influences employees' career capabilities. Regarding service and technical capabilities, integrating job crafting theory and conservation of resources theory, Mo et al. (2024) found that AI awareness can positively enhance employees' service performance adaptability and technologically-environmental mastery through the mediating pathways of defensive and promotive job crafting. Concurrently, servant leadership moderates the relationship between AI awareness and defensive job crafting but moderates its relationship with promotive job crafting. At the vocational capability level, scholars combining

career theory, fit theory, and organizational commitment theory discovered that the higher the organizational commitment, the more pronounced the potential inhibitory effect of AI awareness on vocational capability (Kong et al., 2021). In the creativity dimension, based on job regulation theory and locus of control theory, it was found that AI and robotics awareness can promote employee creativity through the mediating roles of active learning and task crafting (Wang et al., 2022).Fifth, AI awareness impacts workplace well-being. Based on the job demands-resources model, Arboh et al. (2025) revealed that AI awareness positively influences workplace well-being, with informal learning behavior serving as a mediator between them. Moreover, the stronger the learning orientation, the more significant the effect of AI awareness on enhancing well-being through informal learning behavior.

In the next place, regarding employee behaviors, scholars have primarily investigated the impact of AI awareness on behaviors targeted at the organization, the family, and customers.First, AI awareness influences organization-oriented behaviors. In terms of proactive behaviors, He et al. (2025), drawing on the transactional theory of stress, found that AI awareness negatively affects both promotive voice and prohibitive voice. This process is mediated by voice efficacy and job insecurity. Moreover, trait competitiveness moderates the relationship between AI awareness and voice efficacy but has no significant moderating effect on the relationship between AI awareness and job insecurity. Based on conservation of resources theory, Lin and He (2024) further confirmed that AI awareness positively influences voice behavior through the mediating roles of learning orientation and supervisor-subordinate guanxi. In addition, based on social cognitive theory and the job demands-resources model, Zhou and Lyu (2025) found that leader AI awareness positively affects employee voice behavior. Specifically, leader AI awareness increases both employees' job insecurity and intrinsic motivation and positively moderates the effect of AI self-efficacy on voice behavior. While job insecurity negatively affects voice behavior, intrinsic motivation has a positive impact.Based on protection motivation theory, Yin et al. (2024) found that AI awareness interacts with change-oriented leadership to influence employee-AI collaboration through approach and avoidance motivations. When AI awareness is perceived, strong change-oriented leadership tends to trigger employees' approach motivation, which promotes collaboration, whereas weak change-oriented leadership induces avoidance motivation, reducing collaboration. Furthermore, drawing on self-efficacy theory and the stimulus-organism-response model, Yan and Teng (2025) identified dual pathways in the relationship between AI awareness and organizational citizenship behavior. Performance self-efficacy mediates the positive effect of AI awareness, while creative self-efficacy mediates a negative effect. They also found that a stronger climate for innovation enhances the positive influence of AI awareness on organizational citizenship behavior via performance self-efficacy.In terms of unethical behaviors, combining conservation of resources theory and cognitive appraisal theory of stress, Zhou and Yi (2025) found that AI awareness positively affects unethical pro-organizational behavior through the mediation of job insecurity.Based on conservation of resources theory, Teng et al. (2024) revealed that AI awareness positively influences work withdrawal through a sequential mediation of negative work-related rumination and emotional exhaustion. Scholars also confirmed that AI awareness has a positive effect on both physical withdrawal and psychological withdrawal (Nguyen & Nguyen, 2025). Based on conservation of resources theory, Bai et al. (2024) found that psychological contract and emotional exhaustion contribute to the positive effect of AI awareness on counterproductive work behavior.Drawing on the stress-emotion model and social exchange theory, Doan and Nguyen (2025) found that AI awareness positively influences counterproductive work behavior through the mediation of interpersonal conflict. Bai et al. (2024) also found that psychological contract and emotional exhaustion mediate the relationship between AI awareness and counterproductive behavior.Integrating conservation of resources theory and the job demands-resources model, Liu and Cheng (2025) found that AI awareness influences employee resistance to AI through the mediating roles of recovery level and job insecurity. Zhang et al. (2025),

also applying conservation of resources theory, revealed that AI awareness strengthens the positive effect on deviant behavior through ego depletion. Ego depletion mediates the relationship between AI awareness and deviant behavior, and this effect becomes more pronounced when employees have lower tolerance for uncertainty. In addition, based on social exchange theory and moral identity theory, Cheng et al. (2025) found that AI awareness positively influences employee silence through the mediation of psychological contract violation. This effect becomes weaker when moral identity is stronger.Second, AI awareness affects family-oriented behaviors. Based on the stress-emotion model, Zhou et al. (2024) found that AI awareness positively influences work-family conflict, with negative affect mediating this process. They also showed that the effect of AI awareness on work-family conflict through negative emotions becomes weaker when individuals have a higher promotion focus, stronger empowering leadership, or stronger family motivation.Third, AI awareness influences customer-oriented behaviors. Based on the stress-emotion model, Zhou et al. (2024) also discovered that AI awareness positively affects counterproductive work behaviors toward customers, with negative emotions mediating this relationship.similarly, the higher the promotion focus or level of empowering leadership, or the stronger the family motivation, the weaker the promoting effect of AI awareness on counterproductive work behaviors toward customers through negative affect. In addition, based on the job demand-resources model, scholars found a dual mediation mechanism in the relationship between AI awareness and service innovation behavior. Specifically, emotional exhaustion mediates the negative impact of AI awareness, while intrinsic motivation mediates its positive impact. Furthermore, when future orientation is stronger, the negative effect of AI awareness on service innovation through emotional exhaustion becomes weaker, and the positive effect through intrinsic motivation becomes stronger (Liang et al., 2022). Moreover, drawing on the STARA framework and conservation of resources theory, Ma and Ye (2022) found that AI awareness positively influences service sabotage through the mediation of organization-based self-esteem. They also found that perceived organizational support moderates this relationship.

Finally, in the context of employee performance, drawing on self-construal theory, Zhao et al. (2025) found that AI awareness positively influences both in-role and extra-role performance through the mediating effects of perceived overqualification, reflection on AI usage, and dependence on AI usage. They further showed that when the level of employee-AI collaboration is higher, AI awareness more strongly promotes perceived overqualification, which in turn shapes the pathway through which AI awareness affects performance.

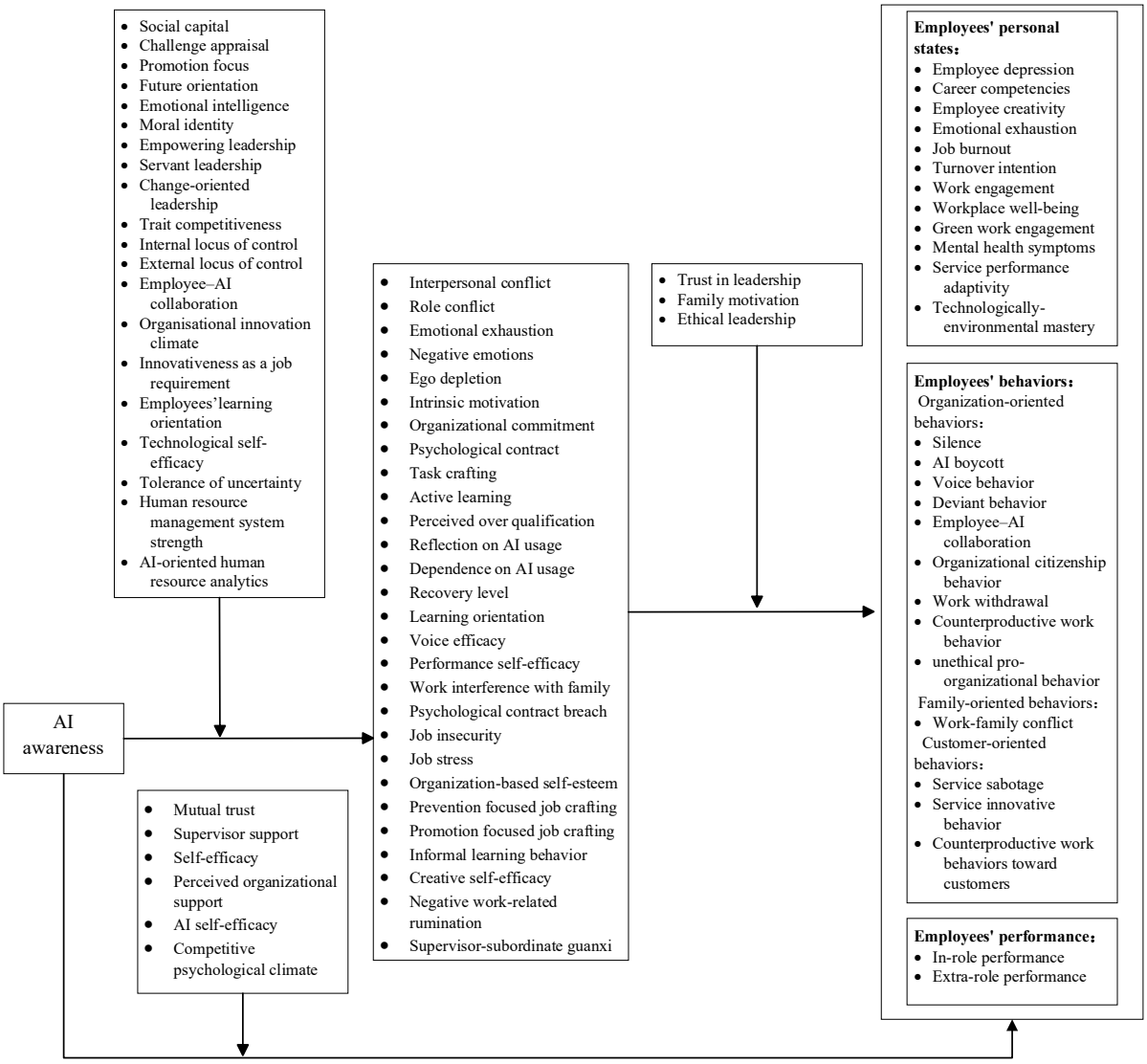


Figure 1: Research of AI awareness

Table 2: The years and countries involved in the research on AI awareness

Year	Country	Number	Country	Number	Country	Number	Country	Number
2019	China	1						
2021	USA	1	China	1				
2022	Eurostat	1	Pakistan	1	China	3		
2023	China	1	Korea	1				
2024	Indonesia	1	China	7	Egypt	1	Korea	1
2025	Vietnam	2	Turkish	1	Ghana	1	China	11

Conclusion

This study draws the following main conclusions. First, current research has reached an emerging consensus on the conceptual definition and measurement approaches for AI awareness, with the development of related scales being relatively unified, providing a solid instrumental foundation for subsequent research. Second, the underlying mechanisms that give rise to AI awareness remain underexplored. Existing literature has primarily focused on the outcome variables of AI awareness, while less attention has been given to its antecedents, and a systematic theoretical framework has yet

to be established. Third, further investigation is needed into the consequences of AI awareness. Current studies mostly examine its impact on employees' psychological states, behavioral outcomes, and job performance; however, the findings remain inconsistent, and no clear consensus has been reached. Finally, most existing research adopts either a resource-based perspective, such as conservation of resources theory, or a cognitive perspective, such as transactional theory of stress, to explain how AI awareness influences employee reactions. Nonetheless, AI awareness may affect individuals through multiple psychological pathways and social processes, suggesting that its underlying mechanisms could be more complex than currently understood.

Recommendations

Although existing research has made certain progress in the field of AI awareness, several critical limitations remain. In general, the current body of literature still suffers from a lack of systematic theoretical foundations, insufficient exploration of underlying mechanisms, and relatively narrow research perspectives. Therefore, future studies may focus on the following directions.First, more attention should be paid to the antecedents of AI awareness. Existing studies have largely concentrated on its outcome variables (Cheng et al., 2025; Kang et al., 2024), while investigations into its formation mechanisms remain limited. In particular, there is a notable absence of multi-level and systematic theoretical models. Future research should devote greater effort to understanding the developmental pathways of AI awareness by incorporating factors at the individual, team, and organizational levels. At the individual level, variables such as employees' professional identity, growth mindset, and satisfaction of autonomy needs may exert significant influence on their perception of AI. At the team level, elements including team climate, culture of change, leadership style, and communication approaches regarding technology may also shape employees' awareness of AI. At the organizational level, the degree of digital transformation, frequency of technological change, clarity of technology use norms, and availability of training programs are all likely to either suppress or amplify employees' AI awareness. Hence, future studies are encouraged to integrate insights from multiple theoretical perspectives and to construct multi-level models that systematically uncover the antecedents and interaction mechanisms of AI awareness. Such efforts would contribute to building a more solid theoretical foundation for understanding its formation process.

Second, it is essential to further explore the functional effects of AI awareness. Existing research on the consequences of AI awareness has primarily focused on employees' psychological states, such as work engagement and turnover intention, as well as behavioral responses, including voice behavior, silence, resistance to AI, and service sabotage, in addition to job performance (Cheng et al., 2025; Kang et al., 2024; Liu & Cheng, 2025; Lin & He, 2024; Ma & Ye, 2022). However, the findings across these studies remain somewhat inconsistent. While some scholars regard AI awareness as a threat that tends to trigger negative employee behaviors, others highlight its positive potential to stimulate employees' sense of urgency and motivation for action, which may, in turn, foster proactive behaviors. Therefore, future research should aim to uncover the double-edged effects of AI awareness and to clarify the conditions and boundaries under which it exerts positive or negative influences. This can be achieved by incorporating moderating variables such as employees' self-efficacy, perceptions of organizational justice, career security, and job autonomy, in order to identify when and under what circumstances AI awareness produces beneficial or detrimental outcomes. In addition, future studies may move beyond the current focus on linear relationships between AI awareness and its consequences to examine the possibility of nonlinear dynamics. For instance, to a certain extent, AI awareness may enhance employees' sense of workplace self-preservation by encouraging skill development and improved service behavior; however, when the level of AI awareness is either too low or excessively high, employees may instead adopt passive coping strategies.

Third, broaden the theoretical perspectives on AI awareness's influence mechanisms to uncover its multi-pathway processes. Current research on the mechanisms through which AI awareness exerts its effects predominantly employs theoretical lenses such as conservation of resources theory and the Transactional Theory of Stress, explaining employee behavioral reactions from perspectives of resource flow or cognitive appraisal (Liu & Cheng, 2025; Yan et al., 2025; Zhang et al., 2025; Zhou & Yi, 2025). AI awareness, as a subjective perception concerning the potential for future career development and role substitution, involves complex and diverse influence mechanisms and pathways; it is more likely the result of intertwined multiple psychological pathways and social processes rather than a singular linear process. Future research should introduce more diverse theoretical perspectives: for instance, self-determination theory could explain how AI awareness influences employees' intrinsic motivation and behavioral regulation; social identity theory could reveal how AI awareness alters employees' cognitions regarding professional identity and group belonging, thereby affecting behavioral performance; affective events theory would help understand how emotional experiences triggered by AI awareness influence employees' momentary decision-making and behavioral execution. Future studies should adopt multi-theoretical integration to more comprehensively unveil the psychological transmission pathways of AI awareness, thereby fostering deeper theoretical development in this field.

References

Arboh, F., Zhu, X., Atingabili, S., Yeboah, E., & Drokow, E. K. (2025). From fear to empowerment: The impact of employees AI awareness on workplace well-being–a new insight from the JD-R model. *Journal of Health Organization and Management*. Advance online publication.

Arias-Pérez, J., & Vélez-Jaramillo, J. (2022a). Understanding knowledge hiding under technological turbulence caused by artificial intelligence and robotics. *Journal of Knowledge Management*, *26*(6), 1476-1491.

Arias-Pérez, J., & Vélez-Jaramillo, J. (2022b). Ignoring the three-way interaction of digital orientation, not-invented-here syndrome and employee's artificial intelligence awareness in digital innovation performance: A recipe for failure. *Technological Forecasting and Social Change*, *174*, Article 121305.

Armstrong-Stassen, M. (2001). Reactions of older employees to organizational downsizing: The role of gender, job level, and time. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *56*(4), P234-P243.

Bai, S., Zhang, X., Yu, D., & Yao, J. (2024). Assist me or replace me? Uncovering the influence of AI awareness on employees' counterproductive work behaviors. *Frontiers in Public Health*, *12*, Article 1449561.

Bakir, S., Dogru, T., Bilgihan, A., & Ayoun, B. (2025). AI awareness and employee-related outcomes: A systematic review of the hospitality literature and a framework for future research. *International Journal of Hospitality Management*, *117*, Article 103654.

Barnes, E., & Hutson, J. (2024). AI and the cognitive sense of self. *Journal of Intelligent Communication*, *3*(1), 35-51.

Brougham, D., & Haar, J. (2018). Smart technology, artificial intelligence, robotics, and algorithms (STARA): Employees' perceptions of our future workplace. *Journal of Management & Organization*, *24*(2), 239-257.

Cheng, M., Zhang, L., & Wang, H. (2025). The effect of artificial intelligence awareness on frontline service employees' silence: The roles of psychological contract breach and moral identity. *International Journal of Contemporary Hospitality Management*, *37*(5), 1845-1861.

Doan, X. H., & Nguyen, T. P. L. (2025). Moderated mediation model of relationship between artificial intelligence awareness and counterproductive work behavior, turnover intention. *Journal of Organizational Change Management*, *38*(2), 501-519.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerization? *Technological Forecasting and Social Change*, *114*, 254-280.

Hassan, K., Thakur, A. K., Singh, G., Singh, J., Gupta, L. R., & Singh, R. (2024). Application of artificial intelligence in aerospace engineering and its future directions: A systematic quantitative literature review. *Archives of Computational Methods in Engineering*, *31*(7), 4031-4086.

He, C., Teng, R., & Song, J. (2024). Linking employees' challenge-hindrance appraisals toward AI to service performance: the influences of job crafting, job insecurity and AI knowledge. *International Journal of Contemporary Hospitality Management*, *36*(3), 975-994.

He, C., Xiong, H., Cai, W., & Song, J. (2025). How does AI awareness affect employees' voice behavior in the service industry? A transactional theory of stress perspective. *International Journal of Contemporary Hospitality Management*. Advance online publication.

Kang, D. Y., Hur, W. M., & Shin, Y. (2023). Smart technology and service employees' job crafting: Relationship between STARA awareness, performance pressure, receiving and giving help, and job crafting. *Journal of Retailing and Consumer Services*, *73*, Article 103282.

Kang, J., Shin, H., & Kang, C. (2024). Hospitality labor leakage and dynamic turnover behaviors in the age of artificial intelligence and robotics. *Journal of Hospitality and Tourism Technology*, *15*(5), 916-933.

Karalis, V. D. (2024). The integration of artificial intelligence into clinical practice. *Applied Biosciences*, *3*(1), 14-44.

Khairy, H. A., Ahmed, M., Asiri, A., Gazzawe, F., Abdel Fatah, M. A., Ahmad, N., Qahmash, A., & Agina, M. F. (2024). Catalyzing green work engagement in hotel businesses: Leveraging artificial intelligence. *Sustainability*, *16*(16), Article 7102.

Khaliq, A., Waqas, A., Nisar, Q. A., Haider, S., & Asghar, Z. (2022). Application of AI and robotics in hospitality sector: A resource gain and resource loss perspective. *Technology in Society*, *68*, Article 101807.

Kong, H., Yuan, Y., Baruch, Y., Bu, N., Jiang, X., & Wang, K. (2021). Influences of artificial intelligence (AI) awareness on career competency and job burnout. *International Journal of Contemporary Hospitality Management*, *33*(2), 717-734.

Kong, X., Yuan, W., & Ma, C. (2025). Service robots, artificial intelligence awareness, self-efficacy and work engagement. *Management Decision*. Advance online publication.

Koo, B., Curtis, C., & Ryan, B. (2021). Examining the impact of artificial intelligence on hotel employees through job insecurity perspectives. *International Journal of Hospitality Management*, *95*, Article 102763.

Krabokoukis, T. (2023). Technology tools in hospitality: Mapping the landscape through bibliometric analysis and presentation of a new software solution. *Digital*, *3*(1), 81-96.

Lestari, M. A., Sulistyawati, A. I., & Saif, G. M. S. (2024). Mediating contribution of job crafting to the role of servant leadership and AI in enhancing work engagement. *Advances Educational Innovation*, *1*(1), 26-35.

Li, J. J., Bonn, M. A., & Ye, B. H. (2019). Hotel employee's artificial intelligence and robotics awareness and its impact on turnover intention: The moderating roles of perceived organizational support and competitive psychological climate. *Tourism Management*, *73*, 172-181.

Liang, X., Guo, G., Shu, L., Gong, Q., & Luo, P. (2022). Investigating the double-edged sword effect of AI awareness on employee's service innovative behavior. *Tourism Management*, *92*, Article 104564.

Lin, Q., & He, L. (2024). Does artificial intelligence (AI) awareness affect employees in giving a voice to their organization? A cross-level model. *International Journal of Hospitality Management*, *123*, Article 103947.

Liu, S., & Cheng, P. (2025). The double-edged sword effect of artificial intelligence awareness among hotel employees. *International Journal of Contemporary Hospitality Management*. Advance online publication.

Ma, C., & Ye, J. (2022). Linking artificial intelligence to service sabotage. *The Service Industries Journal*, *42*(13-14), 1054-1074.

Mo, Z., Liu, M. T., & Ma, Y. (2024). How AI awareness can prompt service performance adaptivity and technologically-environmental mastery. *Tourism Management*, *105*, Article 104971.

Moin, M. F., & Zhang, J. Z. (2024). Stemming the tide: Linking AI technology with workers retention. *International Journal of Managing Projects in Business*. Advance online publication.

Nguyen, T. P. L., & Nguyen, D. T. (2025). A moderated mediation model of relationship between artificial intelligence awareness and withdrawal behaviors. *International Journal of Organizational Analysis*. Advance online publication.

Parvez, M. O., Öztüren, A., Cobanoglu, C., Arasli, H., & Eluwole, K. K. (2022). Employees' perception of robots and robot-induced unemployment in hospitality industry under COVID-19 pandemic. *International Journal of Hospitality Management*, *107*, Article 103336.

Rawas, S. (2024). AI: The future of humanity. *Discover Artificial Intelligence*, *4*(1), Article 25.

Shen, L. (2022). The performance evaluation model of hotel green human resources based on internet of things and fuzzy theory. *Mobile Information Systems*, *2022*, Article 4866952.

Shum, C., Min, H. K., Sun, J., Yu, H. C., & He, Z. (2024). Kicking the robots: The roles of transformational leadership and fear on service robot risk awareness and robot abuse relationship. *Journal of Hospitality and Tourism Technology*. Advance online publication.

Teng, R., Zhou, S., Zheng, W., & Ma, C. (2024). Artificial intelligence (AI) awareness and work withdrawal: Evaluating chained mediation through negative work-related rumination and emotional exhaustion. *International Journal of Contemporary Hospitality Management*, *36*(7), 2311-2326.

Wang, H., Zhang, H., Chen, Z., Zhu, J., & Zhang, Y. (2022). Influence of artificial intelligence and robotics awareness on employee creativity in the hotel industry. *Frontiers in Psychology*, *13*, Article 834160.

Xu, G., Xue, M., & Zhao, J. (2023). The association between artificial intelligence awareness and employee depression: The mediating role of emotional exhaustion and the moderating role of perceived organizational support. *International Journal of Environmental Research and Public Health*, *20*(6), Article 5147.

Yan, B., & Teng, Y. (2025). The double-edged sword effect of artificial intelligence awareness on organisational citizenship behaviour: a study based on knowledge workers. *Behaviour & Information Technology*. Advance online publication.

Yan, J., Xiao, Q., Zhang, S. X., & Guchait, P. (2025). Artificial intelligence in hospitality: A transactional stress perspective on the mental health of hospitality workers. *International Journal of Hospitality Management*, *126*, Article 104029.

Yang, Y., Chi, M., Bi, X., & Xu, Y. (2024). How does the anthropomorphism of service robots impact employees' role service behavior in the workplace? *International Journal of Hospitality Management*, *122*, Article 103857.

Yin, Z., Kong, H., Baruch, Y., Decosta, P. L. E., & Yuan, Y. (2024). Interactive effects of AI awareness and change-oriented leadership on employee-AI collaboration: the role of approach and avoidance motivation. *Tourism Management*, *105*, Article 104966.

Zhang, Y., Wang, J., Zhang, J., & Wang, Y. (2025). To be right on the button: How and when hotel frontline service employees' AI awareness influences deviant behavior. *International Journal of Hospitality Management*, *126*, Article 104090.

Zhao, H., Ye, L., Guo, M., & Deng, Y. (2025). Reflection or dependence: How AI awareness affects employees' in-role and extra-role performance? *Behavioral Sciences*, *15*(2), Article 128.

Zhao, J., Hu, E., Han, M., Jiang, K., & Shan, H. (2023). That honey, my arsenic: The influence of advanced technologies on service employees' organizational deviance. *Journal of Retailing and Consumer Services*, *75*, Article 103490.

Zheng, J., & Zhang, T. (2025). Association between AI awareness and emotional exhaustion: the serial mediation of job insecurity and work interference with family. *Behavioral Sciences*, *15*(4), Article 401.

Zhou, S., & Yi, N. (2025). How and when does AI awareness affect hotel frontline employee's unethical pro-organizational behavior? Insights from two longitudinal studies. *International Journal of Hospitality Management*, *130*, Article 104267.

Zhou, S., Yi, N., Rasiah, R., Zhao, H., & Mo, Z. (2024). An empirical study on the dark side of service employees' AI awareness: behavioral responses, emotional mechanisms, and mitigating factors. *Journal of Retailing and Consumer Services*, *79*, Article 103869.

Zhou, Y., & Lyu, B. (2025). How does leadership AI awareness shape employee voice behavior? Based on the framework of hindrance and challenge stressors. *Work*, Article 10519815251341816.