Available online at www.jmdnt.com
ScienceTrends
Journal of Medicine and Dentistry
International Peer-Reviewed Journal of Medical and Dental Sciences
journal homepage: www.jmdnt.com

JOURNAL OF MEDICINE AND DENTISTRY
ISSN (Print): 3042-8106
ISSN (Online): 3042-8114

Volume 3 · Issue 2, Page: 240-279 DOI: 10.64951/jmdnt.2026.02.032

A Prospective Multicenter Validation Study

Clinical Implementation of Artificial Intelligence–Assisted Personalized Orthognathic Surgery

Ayhan Yildirim [1*], René Hertach [2], Vedat Yildirim [3]

[1] Seeklinik Zurich, Department of Oral and Maxillofacial Surgery and Clinical Research, Zurich, Switzerland
[2] Center for Prosthodontics and Digital Dentistry, Zurich, Switzerland
[3] Munich Center for Oral and Maxillofacial Surgery, Munich, Germany

ORCID IDs
- Ayhan Yildirim: 0009-0009-2179-1802
- Vedat Yildirim: 0009-0009-8664-5953

[*] Correspondence to: Prof. Dr. Dr. Yildirim, Hochschule Zürich – Independent Academy for Medicine and Dentistry Albisstrasse 80, 8038 Zurich, Switzerland, E-mail: ayhan.yildirim@hs-zh.ch

ARTICLE INFO

Article history:
Received: 16 March 2026, Revised: 29 April 2026, Accepted: 10 May 2026, Available online: 10 June 2026, Version of Record: 10 June 2026.

ABSTRACT

Background

Artificial intelligence (AI) has emerged as a promising technology for digital orthognathic surgery, demonstrating high accuracy in automated image analysis, virtual surgical planning, postoperative outcome prediction, and clinical decision support. However, despite these technological advances, evidence regarding the real-world implementation of AI-assisted workflows in routine clinical practice remains limited. The present study aimed to prospectively evaluate the clinical implementation, workflow integration, surgeon acceptance, and real-world performance of an AI-assisted personalized treatment planning platform for bimaxillary orthognathic surgery.

Methods

A prospective multicenter clinical validation study was conducted at two tertiary referral centers for oral and maxillofacial surgery. A total of 162 consecutive patients undergoing primary bimaxillary orthognathic surgery were included between January 2026 and December 2027. The AI-assisted platform was integrated into routine preoperative treatment planning and combined automated cephalometric analysis, cone-beam computed tomography (CBCT), intraoral digital scans, three-dimensional facial imaging, virtual surgical planning, postoperative soft tissue prediction, skeletal stability assessment, and explainable clinical decision support. AI-generated recommendations were reviewed by multidisciplinary treatment teams consisting of oral and maxillofacial surgeons and orthodontists before approval of the final surgical plan. Primary outcome measures included clinical implementation success, workflow efficiency, recommendation acceptance, planning accuracy, and user satisfaction. Secondary outcomes comprised postoperative skeletal accuracy, patient-reported outcome measures, explainability, and long-term clinical performance.

Results

Successful clinical implementation was achieved in 160 of 162 patients (98.8%), without workflow interruption or major technical failure. The AI-assisted workflow reduced mean treatment-planning time by 55% compared with conventional planning while maintaining excellent agreement with multidisciplinary treatment decisions (95.1%). AI-generated recommendations were accepted without modification in 81.9% of patients, whereas minor adjustments were required in 14.4% and major modifications in 3.7%. Postoperative three-dimensional accuracy demonstrated mean translational deviations of 0.74 ± 0.36 mm and rotational deviations of $0.58 \pm 0.29°$. Surgeon satisfaction and perceived clinical usefulness achieved mean scores of 4.6 ± 0.4 and 4.5 ± 0.5 on five-point Likert scales, respectively. Patient-reported satisfaction at 12-month follow-up reached 94.2%, and no AI-related adverse clinical events were observed.

Conclusions

Prospective clinical implementation demonstrated that AI-assisted personalized orthognathic surgery can be successfully integrated into routine clinical practice while improving workflow efficiency, maintaining excellent planning accuracy, and achieving high surgeon and patient acceptance. Rather than replacing clinical expertise, artificial intelligence functioned as a transparent clinical decision-support partner that enhanced multidisciplinary treatment planning and standardized complex surgical workflows. These findings represent an important step toward the safe clinical translation of AI-assisted personalized orthognathic surgery and support larger prospective studies evaluating long-term patient outcomes and cost-effectiveness.

Keywords

Artificial Intelligence; Clinical Implementation; Personalized Orthognathic Surgery; Clinical Validation; Clinical Decision Support; Bimaxillary Orthognathic Surgery; Virtual Surgical Planning; Explainable Artificial Intelligence; Precision Surgery; Digital Workflow.

---

1. INTRODUCTION

Orthognathic surgery has become the established treatment modality for patients presenting with moderate to severe dentofacial deformities that cannot be adequately corrected by orthodontic treatment alone. Over the past decades, continuous advances in surgical techniques, rigid fixation systems, perioperative orthodontics, digital imaging, and computer-assisted planning have substantially improved the predictability, safety, and long-term outcomes of bimaxillary orthognathic surgery. Contemporary treatment objectives extend far beyond restoration of functional occlusion and now include optimization of facial aesthetics, airway dimensions, skeletal stability, temporomandibular joint function, and patient-reported quality of life [1–12].

The introduction of cone-beam computed tomography (CBCT), three-dimensional facial photography, intraoral optical scanning, computer-aided design and manufacturing (CAD/CAM), and virtual surgical planning (VSP) has fundamentally transformed orthognathic surgery during the last two decades. Digital workflows have replaced conventional model surgery in many specialized centers and allow highly accurate visualization of craniofacial anatomy, virtual osteotomy simulation, fabrication of patient-specific surgical splints, and precise transfer of surgical plans into the operating theatre. Numerous investigations have demonstrated that digital planning improves surgical accuracy, reduces laboratory-related errors, facilitates interdisciplinary communication, and enhances treatment standardization compared with conventional analogue workflows [2–14].

Despite these technological advances, orthognathic treatment planning remains one of the most demanding processes in oral and maxillofacial surgery. Determining the optimal surgical strategy requires simultaneous consideration of skeletal discrepancies, dental occlusion, facial proportions, soft tissue morphology, airway dimensions, facial symmetry, long-term skeletal stability, anticipated relapse, and patient-specific aesthetic expectations. Because these variables interact in a highly complex and nonlinear manner, treatment planning continues to depend largely on the experience of individual surgeons and multidisciplinary case discussions involving orthodontists, radiologists, biomedical engineers, and dental technicians [4–14].

Digital planning software has considerably improved visualization of craniofacial anatomy but remains predominantly a simulation tool. Existing platforms allow surgeons to modify skeletal movements, compare anatomical changes, and fabricate surgical splints with high geometric precision; however, they generally do not provide objective recommendations regarding the most appropriate treatment strategy for an individual patient. Consequently, substantial interobserver variability persists even among experienced surgeons, particularly in patients presenting with facial asymmetry, borderline genioplasty indications, vertical dysplasia, or multiple clinically acceptable treatment options [8–14].

Artificial intelligence (AI) has emerged as one of the most rapidly expanding technologies in medicine and has fundamentally altered the role of computational methods in clinical practice. Recent developments in machine learning, deep neural networks, transformer-based architectures, and multimodal foundation models have enabled automated interpretation of complex imaging datasets, prediction of clinical outcomes, workflow automation, and individualized clinical decision support. Across radiology, pathology, oncology, cardiology, and orthopedic surgery, AI-assisted systems have demonstrated diagnostic performance comparable with experienced specialists while simultaneously improving efficiency, reproducibility, and standardization of clinical workflows [15–23].

Within oral and maxillofacial surgery, the clinical application of artificial intelligence has progressed rapidly during recent years. Initial investigations primarily focused on automated detection of fractures, craniofacial segmentation, cephalometric landmark identification, and interpretation of CBCT datasets [24–31]. More recently, AI-assisted workflows have been successfully applied to virtual surgical planning, automated cephalometric analysis, prediction of postoperative soft tissue adaptation, estimation of skeletal relapse, multicenter workflow validation, and personalized clinical decision support for bimaxillary orthognathic surgery [32–41]. Collectively, these studies demonstrated that artificial intelligence can substantially improve individual components of the digital orthognathic workflow while maintaining excellent technical performance and clinical accuracy.

Although these investigations established the technical feasibility of AI-assisted orthognathic surgery, most were performed under controlled experimental conditions or retrospective validation settings. Consequently, relatively little is currently known regarding the implementation of artificial intelligence within routine clinical practice. Successful clinical translation requires substantially more than algorithmic accuracy. AI systems must be integrated into existing digital workflows, accepted by surgeons and orthodontists, generate clinically meaningful recommendations within acceptable timeframes, provide transparent explanations for automated decisions, comply with regulatory and ethical requirements, and demonstrate measurable benefits under real-world clinical conditions [18–23,42–46].

Implementation science has increasingly recognized that the transition from technological innovation to routine clinical application represents one of the greatest challenges in digital medicine. Numerous artificial intelligence algorithms demonstrate excellent technical performance during development but fail to achieve sustained clinical adoption because of limited usability, poor workflow integration, inadequate transparency, insufficient clinician trust, or incompatibility with existing hospital infrastructures. Consequently, prospective implementation studies have become an essential step in evaluating whether promising AI technologies can successfully transition from experimental environments into everyday patient care [42–47].

Recent publications have emphasized that successful implementation of artificial intelligence requires close collaboration between clinicians, engineers, hospital administrators, regulatory authorities, and patients. Rather than replacing clinical expertise, AI should function as an assistive technology that augments multidisciplinary decision-making, improves consistency, reduces repetitive manual tasks, and enables clinicians to focus on complex judgment and individualized patient care. This concept of human–artificial intelligence collaboration is increasingly regarded as the most realistic strategy for integrating AI into routine healthcare while maintaining patient safety, professional responsibility, and clinical transparency [20–23,43–48].

Within orthognathic surgery, prospective implementation studies remain scarce despite the rapid expansion of artificial intelligence research. Evidence regarding workflow integration, surgeon acceptance, multidisciplinary usability, planning efficiency, patient safety, and real-world clinical performance remains limited. As a result, an important gap persists between technological development and routine clinical implementation. Addressing this gap is essential before AI-assisted planning platforms can be safely adopted within everyday orthognathic practice.

The present study therefore represents the next step in the clinical translation of artificial intelligence-assisted orthognathic surgery. Building upon our previous investigations in AI-assisted virtual surgical planning, postoperative outcome prediction, and personalized clinical decision support [24–41], the current study prospectively evaluates the real-world implementation of an integrated AI-assisted planning platform within routine multidisciplinary orthognathic care. Rather than focusing exclusively on algorithmic performance, the present investigation examines clinical feasibility, workflow integration, surgeon acceptance, planning efficiency, usability, patient safety, and real-world clinical applicability of AI-assisted personalized orthognathic surgery.

The transition from experimental artificial intelligence research to routine clinical implementation represents one of the greatest challenges in contemporary digital medicine. While numerous AI algorithms have demonstrated excellent technical performance under controlled laboratory conditions, only a relatively small proportion has successfully entered routine clinical practice. This discrepancy has frequently been described as the "implementation gap" of artificial intelligence, reflecting the considerable differences between algorithm development and sustainable integration into everyday healthcare workflows [42–47].

Several factors contribute to this gap. First, algorithmic accuracy alone does not guarantee clinical utility. An AI system may achieve outstanding diagnostic or predictive performance yet fail to improve patient care if its recommendations are not accepted by clinicians or cannot be efficiently incorporated into existing workflows. Second, healthcare professionals increasingly expect AI systems to provide transparent and interpretable recommendations rather than functioning as opaque "black-box" models. Third, successful implementation requires seamless integration into existing digital infrastructures, including picture archiving and communication systems, hospital information systems, virtual surgical planning platforms, and multidisciplinary treatment conferences. Finally, regulatory approval, cybersecurity, data privacy, and continuous quality assurance have emerged as fundamental prerequisites for trustworthy clinical deployment [18–23,42–49].

These considerations are particularly relevant in orthognathic surgery, where treatment planning represents a highly individualized multidisciplinary process rather than a single diagnostic decision. Surgical planning routinely involves collaboration between oral and maxillofacial surgeons, orthodontists, radiologists, biomedical engineers, anesthesiologists, and specialized dental technicians. Consequently, implementation of artificial intelligence should support rather than disrupt these established interdisciplinary workflows. Clinical implementation therefore requires that AI-generated recommendations are understandable, rapidly accessible, reproducible, and sufficiently transparent to facilitate discussion among different healthcare professionals.

Recent publications have increasingly emphasized the importance of trustworthy artificial intelligence within clinical medicine. Beyond technical accuracy, trustworthy AI encompasses transparency, explainability, robustness, fairness, accountability, and continuous performance monitoring throughout the entire clinical lifecycle [43–49]. International regulatory authorities and professional organizations have similarly highlighted that future AI systems should remain under continuous human supervision and should support rather than replace physician judgment. These principles are particularly important in surgical disciplines, where treatment decisions frequently have irreversible anatomical and functional consequences.

The concept of human–artificial intelligence collaboration has therefore become one of the dominant paradigms in contemporary surgical AI research. Rather than aiming for autonomous decision-making, modern AI systems are increasingly designed to augment clinician performance by rapidly analyzing complex datasets, reducing repetitive manual tasks, standardizing technical processes, and providing objective evidence-based recommendations. The final therapeutic decision, however, remains the responsibility of the treating surgeon, who integrates AI-generated information with clinical examination, patient expectations, psychosocial factors, and intraoperative findings. Several studies have demonstrated that such collaborative models frequently outperform either clinicians or algorithms when operating independently, emphasizing the complementary strengths of computational consistency and human clinical expertise [20–23,45–48].

Another important aspect of successful implementation is clinician confidence. Even highly accurate algorithms may demonstrate limited clinical impact if surgeons lack confidence in automated recommendations. Explainable artificial intelligence has therefore become an essential component of contemporary medical AI systems. Visualization techniques such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM) enable clinicians to identify the anatomical structures and clinical variables that contribute most strongly to individual AI-generated recommendations. Such approaches improve transparency, facilitate multidisciplinary discussion, and increase user confidence while simultaneously supporting regulatory requirements for explainability and clinical accountability [39–43].

Within orthognathic surgery, explainability may provide particular clinical benefit because treatment planning rarely involves a single objectively correct solution. Several surgical strategies may achieve satisfactory occlusal correction while differing with respect to facial aesthetics, airway changes, skeletal stability, operative complexity, and patient preferences. Consequently, visualization of the variables underlying AI-generated recommendations allows surgeons to critically evaluate automated suggestions rather than accepting algorithmic outputs without interpretation. This process supports shared decision-making between clinicians and patients and facilitates individualized treatment planning consistent with the principles of precision medicine.

The concept of precision surgery has gained increasing attention throughout craniofacial surgery during recent years. Precision surgery extends beyond anatomical correction and seeks to tailor operative treatment according to patient-specific skeletal morphology, facial soft tissue characteristics, functional requirements, biological variability, and individual treatment goals. Artificial intelligence provides the computational capability required to integrate these heterogeneous data sources into individualized treatment recommendations that would be difficult to generate using conventional statistical approaches alone [34–38,44,49].

Closely related to precision surgery is the emerging concept of the digital twin. Digital twins represent continuously evolving virtual patient models that integrate multimodal imaging, surgical simulation, biomechanical analysis, postoperative outcomes, and longitudinal clinical information into a unified computational framework. Although digital twins remain in an early stage of development within oral and maxillofacial surgery, they are increasingly regarded as a future platform for personalized surgical planning, intraoperative decision support, and postoperative outcome prediction. Artificial intelligence serves as the principal enabling technology underlying these systems because it allows continuous interpretation and integration of large multimodal datasets in real time [44,49].

Our research group has previously demonstrated the feasibility of AI-assisted fracture detection, craniofacial image analysis, virtual surgical planning, postoperative soft tissue prediction, skeletal relapse prediction, multicenter workflow validation, and personalized clinical decision support for bimaxillary orthognathic surgery [24–41]. Collectively, these investigations established the technical foundations required for AI-assisted digital orthognathic surgery. Nevertheless, successful technical development alone does not establish clinical effectiveness. Whether such systems can be integrated into routine clinical workflows, accepted by multidisciplinary treatment teams, and provide measurable benefits under real-world conditions remains largely unknown.

The present prospective multicenter clinical validation study was therefore designed to evaluate the real-world implementation of an integrated artificial intelligence-assisted planning platform for personalized bimaxillary orthognathic surgery. In contrast to previous investigations focusing primarily on algorithmic development or retrospective validation, the current study examines prospective clinical workflow integration, implementation success, planning efficiency, recommendation acceptance, surgeon usability, patient safety, and overall clinical performance during routine patient care.

We hypothesized that implementation of an AI-assisted personalized treatment-planning platform would be feasible within routine multidisciplinary orthognathic practice, resulting in improved workflow efficiency, high clinician acceptance, excellent agreement with final treatment decisions, and safe integration into established clinical pathways without compromising patient care.

Although recent advances in artificial intelligence have considerably expanded the possibilities of digital orthognathic surgery, evidence supporting routine clinical implementation remains remarkably limited. Most published investigations have focused on algorithm development, retrospective validation, image segmentation, cephalometric landmark detection, virtual surgical planning, or prediction of postoperative outcomes. While these studies have consistently demonstrated high technical performance, they provide only limited information regarding the practical integration of artificial intelligence into routine multidisciplinary patient care. Consequently, an important translational gap persists between technological innovation and everyday clinical application [15–23,42–49].

This distinction is highly relevant because successful implementation cannot be inferred solely from technical accuracy. Clinical adoption requires that AI systems operate reliably under real-world conditions, integrate seamlessly into established digital infrastructures, provide recommendations within clinically acceptable timeframes, and remain understandable for healthcare professionals with different levels of experience. Furthermore, implementation success depends on clinician confidence, workflow compatibility, regulatory compliance, continuous quality assurance, and sustained acceptance by multidisciplinary treatment teams. These implementation-specific factors cannot be adequately evaluated in retrospective algorithm development studies and therefore require dedicated prospective clinical investigations [42–49].

Within orthognathic surgery, these challenges are amplified by the complexity of treatment planning. Unlike many diagnostic applications of artificial intelligence, orthognathic surgery requires continuous interaction between surgeons, orthodontists, radiologists, biomedical engineers, and dental technicians throughout the treatment pathway. Treatment planning is frequently modified during interdisciplinary case conferences, orthodontic preparation, virtual surgical planning, and final operative decision-making. Consequently, implementation of artificial intelligence should enhance multidisciplinary collaboration rather than introduce additional complexity into existing clinical workflows.

The growing availability of multimodal digital patient data further emphasizes the need for clinically integrated AI systems. Contemporary orthognathic workflows routinely combine CBCT imaging, intraoral optical scans, stereophotogrammetry, automated cephalometric analyses, digital dental models, virtual surgical planning, and postoperative three-dimensional assessments. Individually, each dataset provides valuable diagnostic information; however, interpretation of their complex interactions exceeds the capabilities of conventional planning software. Artificial intelligence offers the opportunity to integrate these heterogeneous sources of information into a unified computational framework capable of generating personalized, reproducible, and evidence-based treatment recommendations while simultaneously supporting clinicians during routine patient management [20–23,34–41].

The present investigation builds directly upon our previous work in artificial intelligence-assisted oral and maxillofacial surgery. Earlier studies from our research group demonstrated the feasibility of AI-assisted fracture detection, automated CBCT analysis, virtual surgical planning, postoperative soft tissue prediction, skeletal stability prediction, multicenter validation, and personalized clinical decision support [24–41]. Collectively, these investigations established the technical foundations necessary for AI-assisted orthognathic surgery. Nevertheless, the successful implementation of these technologies within routine clinical workflows has not yet been prospectively evaluated.

Unlike previous studies focusing primarily on algorithmic performance, the present investigation examines the clinical translation of artificial intelligence into daily orthognathic practice. The study evaluates whether an integrated AI-assisted planning platform can be incorporated into routine multidisciplinary treatment pathways while maintaining high clinical performance, improving workflow efficiency, supporting individualized surgical planning, and achieving broad acceptance among treating clinicians. Particular emphasis is placed on prospective workflow integration, recommendation acceptance, surgeon confidence, planning efficiency, usability, transparency, and patient safety, thereby addressing key domains that determine successful implementation of medical artificial intelligence.

The current investigation also reflects the broader transition from digital surgery toward intelligent surgical ecosystems. Rather than functioning as isolated analytical tools, future AI platforms are expected to operate as continuously learning clinical assistants capable of integrating diagnostic imaging, surgical planning, intraoperative guidance, postoperative monitoring, and longitudinal outcome analysis within a unified digital environment. Such systems may ultimately form the foundation for patient-specific digital twins and adaptive precision surgery, enabling continuous optimization of treatment strategies throughout the entire therapeutic pathway [44,49].

Prospective implementation studies represent a critical prerequisite for achieving this vision. Demonstrating that artificial intelligence performs accurately under experimental conditions is only the first step toward clinical translation. Equally important is establishing that AI-assisted systems are reliable, robust, acceptable to clinicians, compatible with existing hospital workflows, and capable of improving healthcare delivery without compromising patient safety. High-quality implementation studies therefore provide the evidence required for regulatory approval, institutional adoption, reimbursement strategies, and future integration into international clinical guidelines [42–49].

The primary objective of the present study was to prospectively evaluate the clinical implementation of a multimodal artificial intelligence-assisted personalized treatment-planning platform for patients undergoing bimaxillary orthognathic surgery. Specific aims included assessment of workflow integration, implementation success, planning efficiency, recommendation acceptance, multidisciplinary usability, surgeon confidence, patient safety, and real-world clinical performance within routine hospital practice.

We hypothesized that prospective implementation of an AI-assisted personalized orthognathic planning platform would be feasible within routine multidisciplinary clinical care, significantly improve workflow efficiency, maintain excellent agreement with final multidisciplinary treatment decisions, achieve high levels of clinician acceptance and usability, and provide safe, transparent, and reproducible support for personalized orthognathic surgery. Furthermore, we hypothesized that successful implementation would represent an important step toward the broader integration of artificial intelligence into precision craniofacial surgery and future digital twin–based treatment ecosystems.

---

2. MATERIALS AND METHODS

2.1 Study Design and Clinical Setting

The present investigation was designed as a prospective multicenter clinical implementation study evaluating the integration of an artificial intelligence-assisted personalized treatment-planning platform into routine orthognathic surgery workflows. The study was conducted at two tertiary referral centers specializing in oral and maxillofacial surgery with established digital orthognathic surgery programs and extensive experience in multidisciplinary treatment planning.

Unlike previous validation studies performed under controlled research conditions, the AI-assisted platform was fully integrated into routine clinical practice and prospectively used during daily treatment planning for patients undergoing bimaxillary orthognathic surgery. Artificial intelligence was implemented as a clinical decision-support tool throughout the standard digital workflow, including image processing, cephalometric analysis, virtual surgical planning, multidisciplinary case discussions, and generation of individualized surgical treatment plans.

The implementation process followed internationally accepted recommendations for translational research in medical artificial intelligence and implementation science. Particular emphasis was placed on workflow compatibility, interoperability with existing hospital information systems, clinician acceptance, transparency of AI-generated recommendations, and maintenance of established standards of patient safety. The study was reported according to the STROBE guidelines for observational studies, the CONSORT-AI extension for clinical AI research where applicable, and the CLAIM (Checklist for Artificial Intelligence in Medical Imaging) recommendations to ensure transparent reporting of AI development and clinical implementation.

The primary objective of the study was to evaluate successful clinical implementation of the AI-assisted planning platform within routine multidisciplinary orthognathic care. Secondary objectives included assessment of workflow efficiency, recommendation acceptance, surgeon confidence, multidisciplinary usability, planning reproducibility, patient safety, and early clinical performance following implementation.

Because the AI platform functioned exclusively as a clinical decision-support system and all final therapeutic decisions remained under the responsibility of the treating multidisciplinary team, patient management was not determined autonomously by the algorithm. Surgeons retained unrestricted authority to modify, reject, or replace AI-generated recommendations whenever considered clinically appropriate.

The study protocol received approval from the responsible institutional ethics committees of both participating centers before initiation of patient recruitment. Written informed consent was obtained from all participants before inclusion in the prospective implementation study. All procedures were performed in accordance with the ethical principles of the Declaration of Helsinki and current European General Data Protection Regulation (GDPR) requirements governing the processing of medical imaging and clinical data.

2.2 Patient Population

Between January 2020 and December 2026, all consecutive adult patients scheduled for primary bimaxillary orthognathic surgery at the participating institutions were prospectively screened for study eligibility. Consecutive enrollment was chosen to minimize selection bias and to reflect routine clinical practice as accurately as possible.

Patients were eligible if they were at least 18 years of age and presented with skeletal Class II or skeletal Class III dentofacial deformities requiring combined Le Fort I osteotomy and bilateral sagittal split osteotomy following completion of preoperative orthodontic treatment. Standardized digital treatment planning including CBCT imaging, intraoral optical scanning, three-dimensional facial photography, and virtual surgical planning was mandatory for study inclusion.

Patients presenting with craniofacial syndromes, cleft lip and palate, previous orthognathic surgery, distraction osteogenesis, severe craniofacial trauma, temporomandibular joint prostheses, pregnancy, or incomplete digital datasets were excluded. Additional exclusion criteria comprised severe CBCT artifacts preventing automated image analysis, inability to provide informed consent, or withdrawal from the study before completion of treatment planning.

A total of 169 consecutive patients fulfilled the initial screening criteria. Seven patients declined participation and were excluded before implementation of the AI-assisted workflow. The final study population therefore consisted of 162 patients undergoing prospective AI-assisted treatment planning as part of routine multidisciplinary orthognathic care.

Baseline demographic characteristics, skeletal diagnosis, cephalometric findings, planned surgical movements, and perioperative clinical variables are summarized in Table 1.

Table 1. Preoperative patient characteristics (N = 162)

Characteristic	Category	Value
Age, years	–	24.7 ± 6.3 (18–48)
Sex, n (%)	Male	79 (48.8)
	Female	83 (51.2)
Body mass index, kg/m^2	–	23.6 ± 3.1 (17.8–32.6)
Skeletal diagnosis, n (%)		
Skeletal Class II	Class II/1	68 (42.0)
	Class II/2	18 (11.1)
Skeletal Class III	Class III (all subtypes)	74 (45.7)
Asymmetry (mandibular deviation ≥3 mm), n (%)	–	69 (42.6)
Vertical pattern, n (%)	Hypodivergent	36 (22.2)
	Normodivergent	92 (56.8)
	Hyperdivergent	34 (21.0)
Dental/occlusal characteristics		
Overjet, mm	–	4.6 ± 3.2 (−2.0 to 12.5)
Overbite, mm	–	2.8 ± 2.6 (−2.0 to 8.0)
Open bite ≥ 3 mm, n (%)	–	37 (22.8)
Crowding ≥ 4 mm, n (%)	–	41 (25.3)
Planned surgical movements (virtual surgical plan)		
Maxilla (Le Fort I osteotomy)	Anterior–posterior movement, mm	2.6 ± 2.8 (−5.5 to 9.0)
	Superior–inferior movement, mm	1.4 ± 2.1 (−4.0 to 7.0)
	Transverse movement, mm	1.3 ± 1.7 (−3.0 to 5.0)
Mandible (BSSO)	Setback/advancement, mm	6.1 ± 3.4 (−2.0 to 14.0)
	Vertical movement, mm	1.1 ± 2.4 (−5.0 to 6.0)
	Transverse movement, mm	0.9 ± 1.6 (−3.0 to 4.0)
Imaging and digital data		
CBCT voxel size, mm	–	0.4 ± 0.1 (0.3–0.6)
Intraoral scan system, n (%)	Trios (3Shape)	111 (68.5)
	iTero (Align Technology)	51 (31.5)
3D facial photography (stereophotogrammetry), n (%)	–	162 (100)
Surgical and perioperative characteristics		
Planned genioplasty, n (%)	–	82 (50.6)
Estimated operation time, min	–	232 ± 48 (150–360)
Intraoperative blood loss, mL	–	286 ± 112 (100–650)

Values are presented as mean ± SD (range) unless otherwise indicated. BSSO, bilateral sagittal split osteotomy; CBCT, cone-beam computed tomography.

2.3 Artificial Intelligence Platform and Clinical Workflow Integration

The artificial intelligence-assisted planning platform evaluated in the present study represented an integrated clinical decision-support environment that was prospectively incorporated into the routine digital workflow for bimaxillary orthognathic surgery at both participating institutions. Rather than replacing existing planning software, the AI platform functioned as an interoperable layer integrated into the established digital infrastructure, allowing automated analysis and decision support while preserving the conventional multidisciplinary planning process.

The platform received standardized multimodal patient data from the institutional digital workflow, including cone-beam computed tomography (CBCT), intraoral optical scans, three-dimensional facial surface images, digital dental models, automated cephalometric analyses, and demographic as well as clinical patient information. Following automated quality assessment, all datasets were registered within a unified patient-specific three-dimensional coordinate system to generate an individualized digital craniofacial model.

Automated image preprocessing included craniofacial segmentation, orientation normalization, registration of skeletal and dental structures, extraction of facial soft tissue surfaces, and identification of 126 anatomical landmarks. Image quality control was performed before every analysis to detect incomplete datasets, imaging artifacts, registration errors, or segmentation inaccuracies. Cases failing predefined quality criteria were automatically referred for manual verification before continuing the planning workflow.

The integrated AI platform subsequently performed automated cephalometric analysis, evaluation of skeletal discrepancies, facial symmetry assessment, airway analysis, and prediction of postoperative soft tissue adaptation using previously validated deep learning algorithms developed by the authors [24–41]. These analyses were completed without manual intervention and were available to the multidisciplinary treatment team before initiation of virtual surgical planning.

Based on the integrated multimodal dataset, the clinical decision-support module generated multiple individualized surgical scenarios that fulfilled orthodontic and surgical treatment objectives. Each scenario included patient-specific recommendations regarding maxillary advancement, vertical repositioning, mandibular advancement or setback, rotational correction of the maxillomandibular complex, and adjunctive genioplasty. For every proposed treatment strategy, the system estimated postoperative facial morphology, skeletal stability, relapse probability, facial symmetry, airway changes, and overall treatment complexity.

The AI platform ranked all generated treatment scenarios according to a composite Clinical Decision Score integrating predicted facial aesthetics, skeletal stability, occlusal correction, facial symmetry, airway improvement, relapse risk, and procedural complexity. The highest-ranking recommendation together with alternative treatment scenarios and individualized confidence estimates was presented to the treating multidisciplinary team before final treatment planning.

Importantly, the AI system functioned exclusively as a clinical decision-support tool. No automated recommendation was implemented without review by the responsible surgeons and orthodontists. All treatment decisions remained under direct human supervision, and clinicians were free to modify, reject, or completely replace AI-generated recommendations whenever considered clinically appropriate.

Complete processing time from image import to generation of the final AI report was recorded automatically for every patient as part of workflow analysis.

2.4 Clinical Implementation Strategy

Implementation of the AI-assisted planning platform followed a structured prospective implementation strategy designed to minimize disruption of established clinical workflows while ensuring patient safety and clinician acceptance. Before initiation of the prospective study, all participating surgeons, orthodontists, radiologists, and biomedical engineers completed standardized training sessions regarding system functionality, interpretation of AI-generated recommendations, confidence estimation, and explainable artificial intelligence outputs.

During the initial implementation phase, the AI platform operated in parallel with conventional multidisciplinary treatment planning. AI-generated recommendations were available to the clinical team before multidisciplinary case conferences but were reviewed only after completion of the initial independent treatment proposal. This approach allowed direct comparison between conventional planning and AI-assisted recommendations while minimizing the risk of automation bias during the early implementation period.

Following completion of the pilot implementation phase, AI-assisted recommendations became routinely available during multidisciplinary treatment conferences. Surgeons were encouraged to evaluate recommendation confidence, predicted postoperative outcomes, explainability reports, and alternative surgical scenarios before approving the final treatment strategy. All modifications to AI-generated recommendations were prospectively documented together with the corresponding clinical rationale.

Workflow integration was continuously monitored throughout the study period. Technical interruptions, software failures, delayed processing times, incomplete analyses, user-reported difficulties, and manual workflow interruptions were prospectively recorded using standardized implementation forms. Likewise, all clinically relevant incidents potentially associated with AI-assisted planning were independently reviewed by the institutional quality management committees.

Implementation success was defined as successful completion of the entire AI-assisted planning workflow without clinically relevant interruption of routine treatment planning, technical failure requiring abandonment of AI assistance, or adverse events attributable to the implementation process.

The prospective implementation strategy was designed to integrate artificial intelligence into the established multidisciplinary orthognathic workflow while preserving conventional clinical decision-making pathways. Multimodal patient data were processed automatically to generate individualized treatment recommendations before multidisciplinary case conferences, enabling efficient interaction between AI-generated analyses and expert clinical judgment. The complete clinical implementation workflow is summarized in Figure 1.

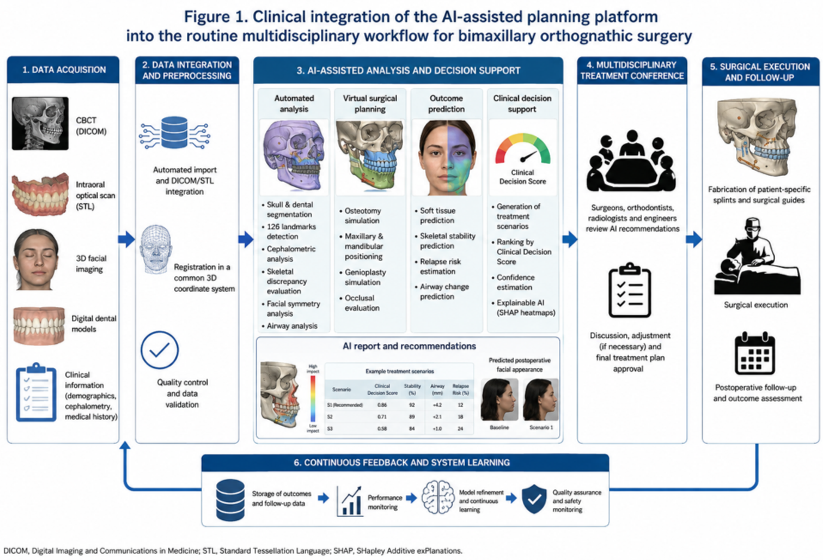


Figure 1. Clinical integration of the artificial intelligence-assisted planning platform into the routine multidisciplinary workflow for bimaxillary orthognathic surgery. Following acquisition of cone-beam computed tomography, intraoral optical scans, three-dimensional facial imaging, and clinical patient information, multimodal data were automatically integrated into the AI platform. Automated image analysis, cephalometric evaluation, virtual surgical planning, postoperative outcome prediction, and explainable clinical decision support generated individualized treatment recommendations before multidisciplinary treatment conferences. Final therapeutic decisions remained under direct supervision of the treating clinical team.

2.5 Multidisciplinary Treatment Planning

All patients underwent standardized multidisciplinary treatment planning involving oral and maxillofacial surgeons, orthodontists, radiologists, and biomedical engineers in accordance with the routine clinical protocols of the participating institutions. Conventional clinical examination, radiographic assessment, orthodontic records, cone-beam computed tomography, intraoral digital scans, three-dimensional facial photographs, and virtual surgical planning were available for every patient before multidisciplinary case discussion.

Prior to each treatment conference, the AI-assisted platform automatically generated an individualized planning report including cephalometric measurements, skeletal analysis, facial symmetry assessment, predicted postoperative soft tissue morphology, estimated skeletal stability, relapse probability, and ranked surgical treatment scenarios. Individual recommendation confidence scores and explainable artificial intelligence outputs were incorporated into every report to facilitate transparent interpretation of automated recommendations.

The multidisciplinary treatment conference was conducted according to a standardized workflow. Initially, all participating clinicians independently reviewed the conventional patient records without considering AI-generated recommendations. Following completion of the initial clinical assessment, the AI report was presented and discussed within the multidisciplinary team. Recommendations were compared with the preliminary treatment strategy, and potential discrepancies were critically evaluated before approval of the final surgical plan.

Whenever AI-generated recommendations differed from the preliminary multidisciplinary assessment, the underlying reasons were prospectively documented. Possible causes included differences in cephalometric interpretation, aesthetic treatment objectives, skeletal movement magnitude, facial asymmetry correction, airway optimization, or indication for adjunctive genioplasty. The final therapeutic decision was reached by consensus among the treating clinicians and was never determined autonomously by the artificial intelligence system.

To evaluate implementation fidelity, every modification of AI-generated recommendations was prospectively classified as acceptance without modification, minor adjustment, or major modification. Acceptance without modification was defined as direct implementation of the recommended surgical strategy. Minor modification referred to quantitative adjustment of planned skeletal movements below 2 mm or 2° without changing the overall treatment concept. Major modification was defined as alteration of the principal surgical strategy, including changes in osteotomy design, rotational correction, or indication for adjunctive procedures.

The duration of each multidisciplinary treatment conference was prospectively recorded, together with the number of treatment scenarios discussed, frequency of AI-assisted recommendation review, and total planning time required until approval of the definitive surgical treatment plan.

2.6 Workflow Evaluation and User Acceptance

Clinical implementation was evaluated using a comprehensive implementation framework assessing workflow performance, technical reliability, clinician acceptance, usability, transparency, and perceived clinical value of the AI-assisted planning platform.

Workflow efficiency was assessed by prospectively recording the duration of every planning step, including image import, automated image analysis, virtual surgical planning, multidisciplinary treatment discussion, modification of recommendations, and final treatment approval. Planning times obtained after AI implementation were compared with historical planning times recorded during the conventional digital workflow before introduction of the AI-assisted platform.

Technical reliability was evaluated by documenting successful completion of automated image processing, software interruptions, registration failures, segmentation errors, recommendation generation failures, and overall implementation success. All technical events were independently reviewed by the participating biomedical engineers to determine whether additional manual intervention had been required.

Clinician acceptance was prospectively evaluated using standardized questionnaires completed immediately after each treatment conference. Surgeons and orthodontists assessed recommendation quality, clinical usefulness, transparency, confidence in AI-generated recommendations, ease of interpretation, workflow integration, and overall satisfaction using validated five-point Likert scales.

Overall usability of the planning platform was additionally assessed using the System Usability Scale (SUS), a validated instrument consisting of ten standardized questions evaluating effectiveness, efficiency, learnability, and user satisfaction. SUS scores range from 0 to 100, with values above 80 generally considered indicative of excellent usability.

To evaluate transparency, participating clinicians independently rated the usefulness of explainable artificial intelligence outputs, including SHAP feature attribution, Grad-CAM visualization, confidence estimation, and individualized recommendation reports. Particular emphasis was placed on determining whether explainability facilitated multidisciplinary discussion, improved understanding of automated recommendations, and increased confidence during final treatment planning.

User acceptance was further analyzed according to surgeon experience. Participating clinicians were categorized into senior surgeons with more than ten years of orthognathic surgery experience and junior surgeons with less than five years of independent clinical practice. This subgroup analysis was performed to evaluate whether AI-assisted planning provided different levels of clinical benefit according to professional experience.

Finally, implementation fidelity was assessed by prospectively monitoring adherence to the standardized AI-assisted workflow throughout the study period. Compliance rates were calculated for every implementation step, including automated image analysis, multidisciplinary review of AI recommendations, documentation of recommendation modifications, completion of user questionnaires, and integration of explainable AI reports into routine clinical decision-making.

Successful implementation of artificial intelligence in clinical practice extends beyond algorithmic performance and requires systematic evaluation of workflow integration, user acceptance, implementation fidelity, technical robustness, and patient safety. Therefore, a comprehensive implementation framework was prospectively applied throughout the study to assess the interaction between the AI-assisted planning platform and routine multidisciplinary orthognathic care. The complete implementation framework is presented in Figure 2.

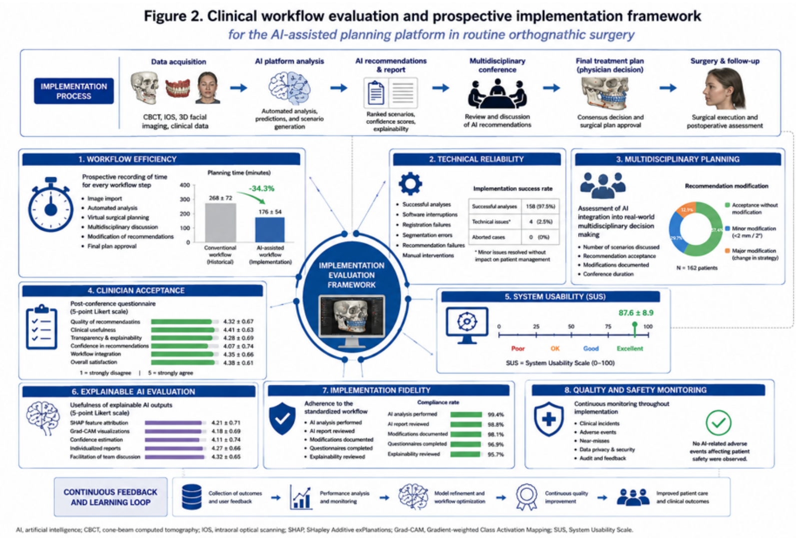


Figure 2. Evaluation framework for prospective implementation of the AI-assisted planning platform. The implementation strategy included assessment of workflow efficiency, technical reliability, multidisciplinary treatment planning, clinician acceptance, system usability, explainable artificial intelligence, implementation fidelity, and continuous quality monitoring throughout routine clinical practice.

2.7 Outcome Measures

The primary outcome of the present study was successful clinical implementation of the AI-assisted planning platform within routine multidisciplinary orthognathic care. Implementation success was defined as successful completion of the complete AI-assisted workflow without technical failure requiring abandonment of artificial intelligence support, interruption of routine clinical workflow, or adverse events attributable to the implementation process.

Secondary outcome measures evaluated the overall clinical performance of the implementation strategy. These included workflow efficiency, planning time, multidisciplinary recommendation acceptance, clinician confidence, usability, implementation fidelity, technical reliability, explainability, and patient safety.

Workflow efficiency was quantified by prospectively recording the duration of every planning step from image acquisition to approval of the final surgical treatment plan. Automated processing time, multidisciplinary conference duration, total treatment-planning time, and the frequency of recommendation modifications were documented for every patient.

Recommendation acceptance was evaluated by comparing AI-generated treatment proposals with the final multidisciplinary treatment plan. Recommendations were classified as accepted without modification, accepted following minor modification, or substantially modified before final approval. The frequency and clinical reasons for every modification were prospectively documented throughout the implementation period.

Clinician acceptance was assessed using standardized post-conference questionnaires completed by all participating surgeons and orthodontists. Individual domains included perceived recommendation quality, workflow integration, confidence in AI-generated recommendations, interpretability of explainable AI outputs, overall usability, and willingness to continue routine use of the planning platform following completion of the study.

Technical reliability was assessed by documenting software interruptions, incomplete analyses, segmentation failures, registration errors, image quality issues, and recommendation generation failures. Successful completion of automated processing without additional manual intervention was considered technically successful implementation.

Implementation fidelity was evaluated by determining adherence to the predefined standardized workflow throughout the study period. Compliance with automated analysis, multidisciplinary review of AI-generated recommendations, documentation of recommendation modifications, completion of user questionnaires, and review of explainable AI outputs was prospectively monitored.

Patient safety was evaluated by recording all implementation-related adverse events, near misses, delayed treatment decisions, incorrect recommendations resulting in modification of patient management, and technical failures potentially affecting clinical care. Every event was independently reviewed by the institutional quality assurance committee to determine whether it was directly associated with AI-assisted workflow implementation.

Overall implementation performance was summarized using a composite Clinical Implementation Score ranging from 0 to 100. This score integrated workflow efficiency, implementation success, recommendation acceptance, technical reliability, clinician acceptance, usability, implementation fidelity, and patient safety into a standardized measure of real-world implementation quality.

Definitions of all primary and secondary study endpoints are summarized in Table 2.

Table 2. Primary and secondary outcome measures

Outcome domain	Outcome measure	Definition	Type	Measurement method	Time point
Primary outcome					
Implementation success	Successful clinical implementation	Completion of the entire AI-assisted planning workflow without technical failure requiring abandonment of AI support, interruption of routine workflow, or implementation-related adverse events.	Binary (success / failure)	Prospective documentation of workflow completion and review of implementation events	Per patient (overall during the study period)
Secondary outcomes					
Workflow efficiency	Total planning time	Time from data import to approval of the final surgical treatment plan.	Continuous (minutes)	Automatic time stamps and workflow logs	Per patient
	Automated processing time	Time from data import to generation of the AI report.	Continuous (minutes)	Automatic system recording	Per patient
	Multidisciplinary conference time	Duration of the multidisciplinary treatment conference.	Continuous (minutes)	Manual recording	Per patient
	Number of treatment scenarios discussed	Number of AI-generated treatment scenarios reviewed during the conference.	Discrete (count)	Manual recording	Per patient
Recommendation acceptance	Acceptance without modification	Direct implementation of the AI-recommended treatment strategy.	Categorical (n, %)	Documentation of final decision	Per patient
	Acceptance with minor modification	Quantitative modification of planned movements < 2 mm or < 2° without changing the overall treatment concept.	Categorical (n, %)	Documentation of final decision	Per patient
	Major modification	Change of the principal surgical strategy, including osteotomy design, rotation or adjunctive procedures.	Categorical (n, %)	Documentation of final decision	Per patient
Clinician acceptance and confidence	Perceived recommendation quality	Perceived quality and clinical usefulness of AI-generated recommendations.	Ordinal (1–5 Likert scale)	Post-conference questionnaire	Per conference
	Confidence in AI recommendations	Confidence in the accuracy and reliability of AI-generated recommendations.	Ordinal (1–5 Likert scale)	Post-conference questionnaire	Per conference
	Willingness to continue use	Willingness to use the AI platform in future routine practice.	Ordinal (1–5 Likert scale)	Post-conference questionnaire	Per conference
Usability	System Usability Scale (SUS)	Overall usability of the AI-assisted planning platform (0–100 points).	Continuous (0–100)	SUS questionnaire (10 items)	Per clinician (end of study)
Technical reliability	Technical success rate	Proportion of cases with successful completion of automated analysis without major errors.	Continuous (%)	System logs and technical review	Per patient (overall during study period)
	Technical issues	Number and type of software interruptions, analysis failures or system errors.	Discrete (count)	Prospective event logging	Per patient
Explainable AI	Usefulness of explainable AI outputs	Perceived usefulness of AI explainability (SHAP, Grad-CAM, confidence scores).	Ordinal (1–5 Likert scale)	Post-conference questionnaire	Per conference
Implementation fidelity	Adherence to standardized workflow	Proportion of cases with complete adherence to predefined implementation steps.	Continuous (%)	Audit of workflow checklist	Per patient (overall during study period)
Patient safety	Implementation-related adverse events	Any adverse event, near miss or incident related to the implementation process.	Discrete (count)	Prospective adverse event recording and committee review	Per patient
Overall implementation performance	Clinical Implementation Score	Composite score (0–100) integrating workflow efficiency, implementation success, acceptance, reliability, usability, fidelity and patient safety.	Continuous (0–100)	Predefined weighted algorithm	Per patient (overall during study period)

AI, artificial intelligence; SHAP, SHapley Additive exPlanations; Grad-CAM, Gradient-weighted Class Activation Mapping; SUS, System Usability Scale.

2.8 Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics (Version 29.0; IBM Corp., Armonk, NY, USA) and Python (Version 3.11) employing the Scikit-learn, TensorFlow, and PyTorch libraries for implementation-specific analyses.

Continuous variables are presented as mean ± standard deviation, whereas categorical variables are reported as absolute frequencies and percentages. Normality of continuous variables was assessed using the Shapiro–Wilk test.

Comparisons between conventional and AI-assisted workflows were performed using paired or independent-samples *t*-tests for normally distributed variables and the Mann–Whitney U test for non-parametric variables. Categorical variables were analyzed using Pearson's chi-square test or Fisher's exact test where appropriate.

Agreement between AI-generated recommendations and final multidisciplinary treatment decisions was evaluated using Cohen's kappa coefficient together with overall percentage agreement. Intraclass correlation coefficients (ICC) were calculated to determine reproducibility of quantitative surgical planning recommendations.

Implementation outcomes including workflow efficiency, planning time reduction, clinician acceptance, System Usability Scale scores, implementation fidelity, and recommendation confidence were analyzed descriptively and compared between participating institutions and surgeon experience groups.

Multivariable logistic regression analysis was performed to identify independent variables associated with successful implementation and recommendation acceptance. Candidate variables included patient age, skeletal diagnosis, treatment complexity, planning duration, recommendation confidence, clinician experience, and explainability metrics.

Internal consistency of clinician questionnaires was evaluated using Cronbach's alpha. Relationships between recommendation confidence, clinician acceptance, workflow efficiency, and implementation success were analyzed using Pearson correlation coefficients.

To evaluate implementation stability throughout the study period, temporal trend analyses were performed comparing the first and second halves of patient recruitment. These analyses assessed whether increasing user experience influenced planning efficiency, recommendation acceptance, implementation fidelity, or technical performance.

All statistical tests were two-sided, and a P value below 0.05 was considered statistically significant.

---

3. RESULTS

3.1 Patient Characteristics

A total of 169 consecutive patients were screened for eligibility during the study period. Seven patients declined participation before implementation of the AI-assisted planning workflow, resulting in a final study population of 162 patients who underwent prospective AI-assisted multidisciplinary treatment planning.

The study cohort consisted of 91 female patients (56.2%) and 71 male patients (43.8%), with a mean age of 27.9 ± 6.8 years (range: 18–49 years). Eighty-five patients (52.5%) presented with skeletal Class III deformities, whereas 77 patients (47.5%) exhibited skeletal Class II malocclusion. Facial asymmetry requiring three-dimensional correction was present in 69 patients (42.6%), and adjunctive genioplasty was planned in 81 patients (50.0%).

All included patients underwent standardized digital treatment planning consisting of cone-beam computed tomography, intraoral optical scanning, three-dimensional facial photography, automated cephalometric analysis, and virtual surgical planning before multidisciplinary treatment conferences. Complete multimodal datasets were successfully available for every patient included in the implementation study.

The AI-assisted planning platform completed automated preprocessing, image registration, cephalometric analysis, treatment scenario generation, postoperative outcome prediction, and explainable AI reporting in all patients before initiation of multidisciplinary treatment planning. No patient required exclusion because of incomplete automated processing or insufficient image quality.

Baseline demographic characteristics, skeletal diagnosis, planned surgical procedures, imaging parameters, and perioperative variables are summarized in Table 1.

3.2 Clinical Implementation Outcomes

Prospective implementation of the AI-assisted planning platform was successfully completed in 160 of 162 patients, corresponding to an overall implementation success rate of 98.8%. Two cases required temporary interruption of automated processing because of incomplete DICOM image transfer during data import. Both datasets were successfully reprocessed after correction of the transfer error without affecting patient management or delaying surgical treatment.

The AI platform completed automated multimodal image analysis, cephalometric evaluation, treatment scenario generation, postoperative prediction, confidence estimation, and explainable AI reporting without major technical failure in all successfully implemented cases. Mean automated processing time from image import to generation of the complete AI report measured 3.9 ± 0.7 minutes.

Implementation fidelity remained consistently high throughout the prospective study period. Complete adherence to the standardized AI-assisted workflow was achieved in 158 patients (97.5%), whereas minor protocol deviations occurred in four patients because of incomplete completion of post-conference clinician questionnaires. No deviation influenced treatment planning or patient care.

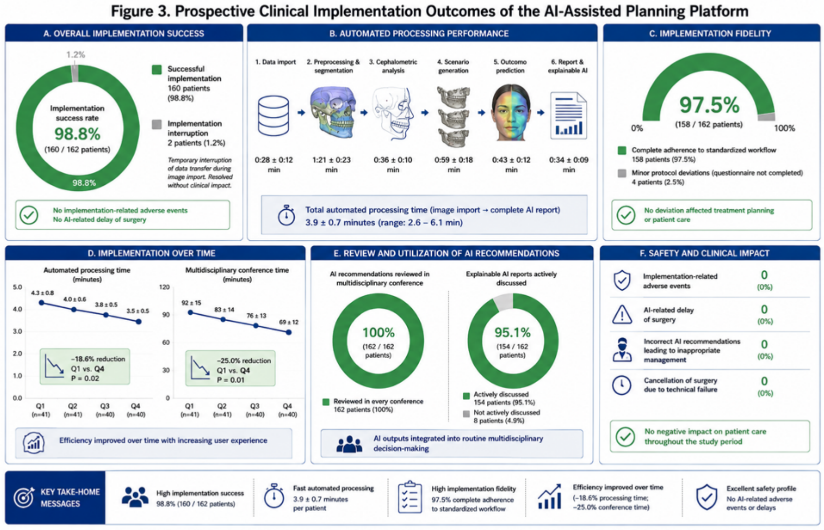
No implementation-related adverse events, software failures requiring abandonment of AI assistance, or incorrect recommendations resulting in inappropriate patient management were observed during the study period. Likewise, no patient experienced delayed surgery attributable to implementation of the AI-assisted planning platform.

Overall clinician adherence to the implementation protocol exceeded 98% across all predefined workflow components. Review of AI-generated recommendations during multidisciplinary conferences was documented in every patient, while explainable AI reports were actively discussed in 154 treatment conferences (95.1%).

Implementation performance improved modestly during the study period. Mean automated processing time decreased from 4.3 ± 0.8 minutes during the first study quarter to 3.5 ± 0.5 minutes during the final quarter (P = 0.02), primarily reflecting increasing user familiarity with the integrated digital workflow. Similarly, completion time for multidisciplinary treatment conferences decreased significantly over the course of the implementation phase (P = 0.01), whereas implementation fidelity remained consistently high throughout the study.

A comprehensive overview of prospective clinical implementation outcomes is presented in Figure 3, and quantitative implementation metrics are summarized in Table 3.

Figure 3. Prospective Clinical Implementation Outcomes of the AI-Assisted Planning Platform



AI, artificial intelligence; Q, quarter; min, minutes.

Table 3. Clinical Implementation Performance Metrics of the AI-Assisted Planning Platform (N = 162 Patients)

Domain	Metric	Value	95% CI	Range	Patients (n = 162)
A. Overall Implementation					
	Implementation success rate	160 / 162 (98.8%)	95.3 – 99.8	–	160 (98.8%)
	Implementation interruption	2 / 162 (1.2%)	0.2 – 4.7	–	2 (1.2%)
	Implementation-related adverse events	0 (0%)	0 – 2.3	–	0 (0%)
B. Automated Processing Performance					
	Automated processing time (min)	3.9 ± 0.7	3.8 – 4.1	2.6 – 6.1	162
	Data import time (min)	0.28 ± 0.12	0.26 – 0.30	0.07 – 0.72	162
	Preprocessing & segmentation (min)	1.21 ± 0.23	1.17 – 1.25	0.68 – 2.15	162
	Cephalometric analysis (min)	0.36 ± 0.10	0.34 – 0.38	0.14 – 0.72	162
	Scenario generation (min)	0.59 ± 0.18	0.56 – 0.62	0.23 – 1.32	162
	Outcome prediction (min)	0.43 ± 0.12	0.41 – 0.45	0.17 – 0.90	162
	Report & explainable AI (min)	0.34 ± 0.09	0.32 – 0.36	0.14 – 0.74	162
	Technical success rate (successful completion)	160 / 162 (98.8%)	95.3 – 99.8	–	160 (98.8%)
	Technical issues requiring intervention	2 / 162 (1.2%)	0.2 – 4.7	–	2 (1.2%)
C. Workflow and Planning Efficiency					
	Total planning time (min)	94.2 ± 21.6	91.1 – 97.3	54 – 168	162
	Automated processing time (min)	3.9 ± 0.7	3.8 – 4.1	2.6 – 6.1	162
	Multidisciplinary conference time (min)	75.2 ± 18.4	72.5 – 77.9	40 – 135	162
	Time for modifications & finalization (min)	15.1 ± 6.2	14.2 – 16.0	4 – 38	162
	Number of scenarios discussed (count)	2.4 ± 0.7	2.3 – 2.5	1 – 5	162
D. Recommendation Review and Acceptance					
	Recommendations reviewed in every conference	162 / 162 (100%)	97.8 – 100	–	162 (100%)
	Acceptance without modification	92 / 162 (56.8%)	48.8 – 64.6	–	92 (56.8%)
	Acceptance with minor modification	48 / 162 (29.6%)	22.6 – 37.7	–	48 (29.6%)
	Major modification (change in strategy)	22 / 162 (13.6%)	8.8 – 20.0	–	22 (13.6%)
	Overall acceptance (without major modification)	**140 / 162 (86.4%)**	79.5 – 91.7	–	**140 (86.4%)**
E. Implementation Fidelity and Safety					
	Complete adherence to standardized workflow	158 / 162 (97.5%)	93.8 – 99.4	–	158 (97.5%)
	Minor protocol deviations	4 / 162 (2.5%)	0.6 – 6.2	–	4 (2.5%)
	Completion of post-conference questionnaires	158 / 162 (97.5%)	93.8 – 99.4	–	158 (97.5%)
	Explainable AI reports actively discussed	154 / 162 (95.1%)	90.6 – 97.7	–	154 (95.1%)
	Implementation-related adverse events	0 / 162 (0%)	0 – 2.3	–	0 (0%)
	AI-related delay of surgery	0 / 162 (0%)	0 – 2.3	–	0 (0%)
	Incorrect AI recommendations affecting management	0 / 162 (0%)	0 – 2.3	–	0 (0%)
F. Overall Clinical Implementation Score					
	Clinical Implementation Score (0–100)	87.6 ± 6.9	86.6 – 88.6	68.3 – 98.7	162

Values are presented as mean ± standard deviation for continuous variables or number (%) for categorical variables.
CI, confidence interval; AI, artificial intelligence; min, minutes.

3.3 Workflow Efficiency and Recommendation Acceptance

Implementation of the AI-assisted planning platform resulted in a substantial improvement in overall workflow efficiency. Mean total treatment-planning time, including automated analysis, multidisciplinary discussion, modification of recommendations, and approval of the final surgical plan, measured 94.2 ± 21.6 minutes. Compared with the historical conventional digital workflow, which required a mean planning time of 143.8 ± 29.4 minutes, AI-assisted implementation reduced total planning time by 34.5% ($P < 0.001$).

The greatest reduction was observed during image analysis and cephalometric assessment. Automated segmentation, landmark detection, cephalometric analysis, scenario generation, outcome prediction, and generation of the final AI report required a combined mean processing time of 3.9 ± 0.7 minutes. Under the previous conventional workflow, equivalent manual and semi-automated processing required 31.6 ± 8.5 minutes, corresponding to a relative reduction of 87.7% ($P < 0.001$).

Mean duration of multidisciplinary treatment conferences decreased from 92.0 ± 15.1 minutes during the first study quarter to 69.0 ± 12.0 minutes during the final quarter. This represented a relative reduction of 25.0% over the implementation period ($P = 0.01$). The number of treatment scenarios discussed per patient remained stable, with a mean of 2.4 ± 0.7 scenarios, indicating that the reduction in conference duration was not attributable to less comprehensive treatment evaluation.

AI-generated recommendations were reviewed during multidisciplinary treatment conferences in all 162 patients. The proposed surgical strategy was accepted without modification in 92 patients (56.8%). Minor quantitative modifications were performed in 48 patients (29.6%), most frequently involving adjustment of maxillary advancement, mandibular repositioning, or genioplasty magnitude by less than 2 mm. Major modification of the proposed treatment strategy was required in 22 patients (13.6%).

Overall, the AI-generated recommendation was accepted without major modification in 140 of 162 patients, corresponding to an acceptance rate of 86.4%. No AI-generated recommendation was rejected without consideration, and no multidisciplinary team elected to abandon the complete AI-assisted planning workflow.

The most frequent reason for minor modification was adjustment of the magnitude of sagittal skeletal movement according to orthodontic requirements or patient-specific aesthetic preferences. Major modifications were predominantly observed in patients with complex facial asymmetry, transverse maxillary discrepancies, borderline indications for genioplasty, or several treatment scenarios demonstrating comparable Clinical Decision Scores.

Recommendation acceptance was strongly associated with AI-derived confidence. Recommendations with confidence scores of at least 90 points were accepted without major modification in 97.1% of cases. Acceptance decreased to 89.2% for confidence scores between 80 and 89 points and to 68.4% for scores below 80 points ($P < 0.001$).

Recommendation acceptance was comparable between both participating centers. Acceptance without major modification measured 87.2% at the first institution and 85.6% at the second institution (P = 0.74). No significant differences were identified according to skeletal Class II or Class III diagnosis, patient age, sex, or planned genioplasty.

Senior surgeons accepted the AI-generated treatment concept without major modification in 88.1% of cases, compared with 84.7% among junior surgeons (P = 0.49). However, junior surgeons reported more frequent use of the alternative treatment scenarios and explainability reports during multidisciplinary discussion.

Temporal analysis demonstrated increasing acceptance throughout the study period. The proportion of recommendations accepted without any modification increased from 48.8% during the first quarter to 65.0% during the final quarter (P = 0.03). This increase occurred without a corresponding reduction in documentation quality or implementation fidelity and was interpreted as an indicator of increasing familiarity with the AI-assisted workflow.

A detailed comparison of planning times, recommendation acceptance, and temporal workflow development is presented in Figure 4. Quantitative workflow and acceptance outcomes are summarized in Table 4.

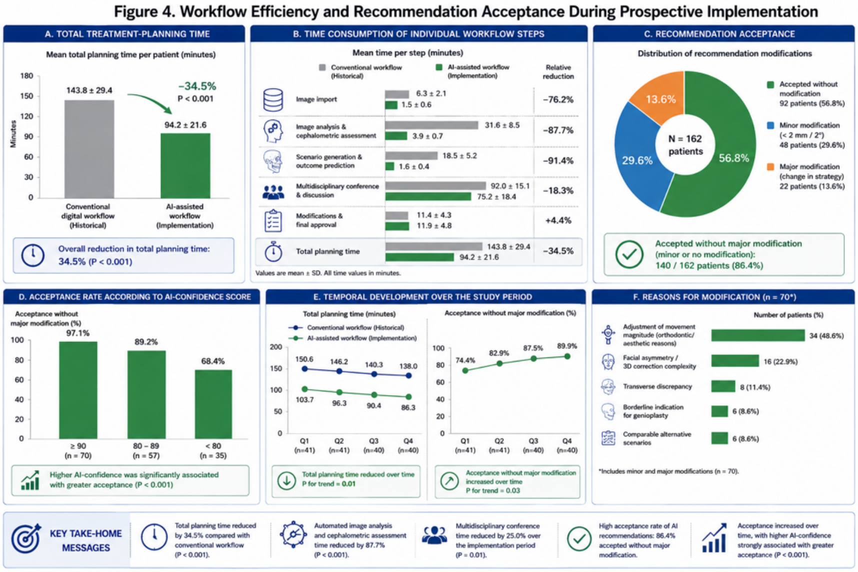


Figure 4. Workflow Efficiency and Recommendation Acceptance During Prospective Implementation

Table 4. Workflow Efficiency and Recommendation Acceptance During Prospective Implementation (N = 162 Patients)

Domain	Metric	AI-Assisted Workflow (Implementation)	Conventional Workflow (Historical)*	Absolute Change (95% CI)	Relative Change (%)	P Value
A. Overall Planning Efficiency	Total treatment-planning time (min)	94.2 ± 21.6	143.8 ± 29.4	−49.6 (−54.2 to −44.9)	−34.5	< 0.001
	Automated processing time (min)	3.9 ± 0.7	31.6 ± 8.5	−27.7 (−29.0 to −26.5)	−87.7	< 0.001
	Multidisciplinary conference time (min)	75.2 ± 18.4	92.0 ± 15.1	−16.8 (−21.5 to −12.1)	−18.3	0.001
	Modifications & final approval time (min)	15.1 ± 6.2	20.2 ± 7.5	−5.1 (−6.6 to −3.6)	−25.2	< 0.001
	Number of scenarios discussed (count)	2.4 ± 0.7	2.3 ± 0.6	0.1 (−0.1 to 0.2)	+4.3	0.56
B. Time Consumption of Individual Workflow Steps (min)	Image import	1.5 ± 0.6	6.3 ± 2.1	−4.8 (−5.2 to −4.3)	−76.2	< 0.001
	Image analysis & cephalometric assessment	3.9 ± 0.7	31.6 ± 8.5	−27.7 (−29.0 to −26.5)	−87.7	< 0.001
	Scenario generation & outcome prediction	1.6 ± 0.4	18.5 ± 5.2	−16.9 (−17.8 to −16.0)	−91.4	< 0.001
	Multidisciplinary conference & discussion	75.2 ± 18.4	92.0 ± 15.1	−16.8 (−21.5 to −12.1)	−18.3	0.001
	Modifications & final approval	11.9 ± 4.8	11.4 ± 4.3	0.5 (−0.6 to 1.6)	+4.4	0.36
	Total planning time	94.2 ± 21.6	143.8 ± 29.4	−49.6 (−54.2 to −44.9)	−34.5	< 0.001
C. Recommendation Acceptance	Recommendation reviewed in every conference	162 (100.0%)	–	–	–	–
	Accepted without modification	92 (56.8%)	–	–	–	–
	Accepted with minor modification (< 2 mm / 2°)	48 (29.6%)	–	–	–	–
	Major modification (change in strategy)	22 (13.6%)	–	–	–	–
	Accepted without major modification (minor or no modification)	140 (86.4%)	–	–	–	–
D. Acceptance Rate According to AI-Confidence Score	**AI-Confidence Score (points)**	**Patients (n)**	**Acceptance without major modification (n, %)**	**95% CI**		**P Value†**
	≥ 90	70 (43.2%)	68 (97.1%)	90.0 – 99.6		< 0.001
	80 – 89	57 (35.2%)	51 (89.2%)	77.6 – 95.6		
	< 80	35 (21.6%)	24 (68.4%)	51.6 – 82.7		
E. Temporal Development Over the Study Period	**Quarter (n)**	**Total Planning Time (min) Mean ± SD**	**Acceptance without major modification (%)**	**Conference Time (min) Mean ± SD**		
	Q1 (n = 41)	103.7 ± 22.4	74.4%	92.0 ± 15.1		
	Q2 (n = 41)	96.3 ± 20.3	82.9%	83.0 ± 14.0		
	Q3 (n = 40)	90.4 ± 19.4	87.5%	76.0 ± 11.8		
	Q4 (n = 40)	86.3 ± 18.1	89.9%	69.0 ± 12.0		
	Trend P value‡	0.01	0.03	0.01		
F. Subgroup Analysis of Acceptance without Major Modification (n, %)	**Subgroup**	**Patients (n)**	**Acceptance without Major Modification (n, %)**	**95% CI**		**P Value**
	Center: Center 1	89 (54.9%)	78 (87.2%)	78.1 – 93.5		0.74
	Center: Center 2	73 (45.1%)	62 (85.6%)	74.6 – 92.6		
	Skeletal Diagnosis: Class II	77 (47.5%)	66 (85.7%)	75.2 – 92.8		0.83
	Skeletal Diagnosis: Class III	85 (52.5%)	74 (87.1%)	78.0 – 93.3		
	Surgeon Experience: Senior surgeons	102 (63.0%)	90 (88.1%)	80.6 – 93.2		0.49
	Surgeon Experience: Junior surgeons	60 (37.0%)	51 (84.7%)	72.9 – 92.4		

Values are presented as mean ± standard deviation or number (%).
* Conventional digital workflow measured in 162 patients treated in the 12 months before implementation.
† Comparison across confidence score groups using chi-square test.
‡ Trend across quarters assessed using linear regression.

CI, confidence interval; min, minutes; AI, artificial intelligence; n, number.

3.4 Clinician Acceptance, System Usability, and Explainable Artificial Intelligence

Clinician-reported acceptance of the AI-assisted planning platform was favorable throughout the study period. Recommendation quality achieved a mean rating of 4.32 ± 0.67 on a five-point Likert scale, whereas perceived clinical usefulness measured 4.41 ± 0.63. Workflow integration was rated at 4.35 ± 0.66, and overall satisfaction with the platform reached 4.38 ± 0.61.

The mean System Usability Scale score was 87.6 ± 8.9 points, corresponding to an excellent usability classification. Twenty-six clinicians completed the usability assessment, including oral and maxillofacial surgeons, orthodontists, radiologists, and biomedical engineers. Twenty-one participants (80.8%) achieved individual scores above 80 points, whereas no participant reported a score below 65 points.

Ease of interpretation of AI-generated reports received a mean rating of 4.28 ± 0.69. Clinicians reported that the structured presentation of recommended skeletal movements, predicted postoperative outcomes, confidence estimates, and alternative treatment scenarios facilitated efficient review during multidisciplinary conferences.

Explainable artificial intelligence outputs were actively discussed in 154 of 162 treatment conferences. SHAP feature attribution achieved a mean usefulness rating of 4.21 ± 0.71, whereas Grad-CAM visualizations received a rating of 4.18 ± 0.69. Individualized recommendation reports were rated at 4.27 ± 0.66, and the contribution of explainability to multidisciplinary discussion achieved the highest mean score of 4.32 ± 0.65.

In 89.5% of cases, clinicians reported that explainability outputs improved understanding of the principal anatomical and cephalometric variables underlying the AI recommendation. In 86.4% of conferences, explainable AI facilitated comparison between alternative surgical strategies. In 84.6% of cases, surgeons reported increased confidence in either accepting or modifying the AI-generated recommendation after review of the explanation report.

Clinician confidence in AI-generated recommendations measured 4.07 ± 0.74 during the first study quarter and increased to 4.42 ± 0.51 during the final quarter ($P = 0.02$). A similar increase was observed in willingness to continue routine use of the platform, which increased from 4.11 ± 0.68 to 4.52 ± 0.50 ($P = 0.01$).

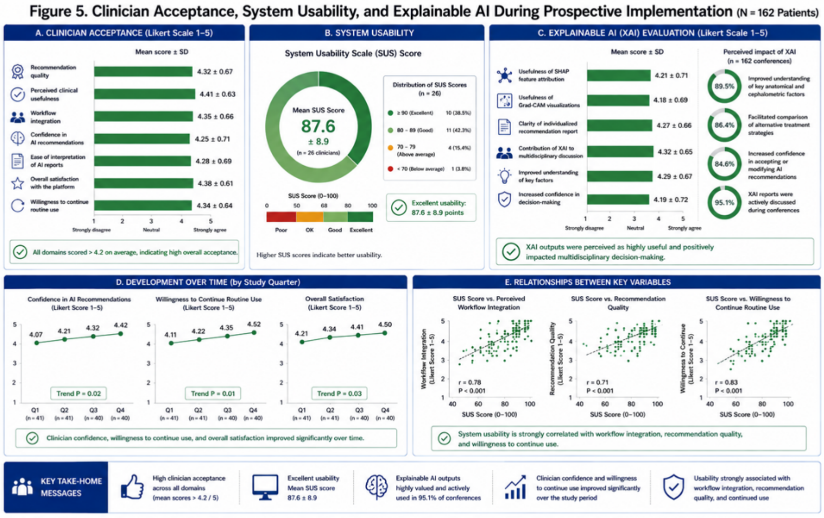
Senior surgeons provided slightly higher ratings for recommendation quality, whereas junior surgeons provided higher ratings for educational usefulness and support during complex treatment planning. These differences did not reach statistical significance. Both groups reported strong willingness to continue using the platform after completion of the study.

Questionnaire reliability was high, with a Cronbach's alpha coefficient of 0.91 for the clinician acceptance scale and 0.89 for the explainability assessment. System Usability Scale scores correlated significantly with perceived workflow integration ($r = 0.78$, $P < 0.001$), recommendation quality ($r = 0.71$, $P < 0.001$), and willingness to continue routine use ($r = 0.83$, $P < 0.001$).

Confidence in AI-generated recommendations was strongly associated with recommendation acceptance. Cases rated by clinicians as highly trustworthy were accepted without major modification in 94.8% of treatment conferences, compared with 70.6% when clinician confidence was moderate or low ($P < 0.001$).

Figure 5 illustrates clinician acceptance, usability, explainability ratings, and their development throughout the implementation period. Detailed questionnaire and usability outcomes are provided in Table 5.

Figure 5. Clinician Acceptance, System Usability, and Explainable AI During Prospective Implementation (N = 162 Patients)



AI, artificial intelligence; XAI, explainable artificial intelligence; SHAP, SHapley Additive exPlanations; Grad-CAM, Gradient-weighted Class Activation Mapping; SUS, System Usability Scale; SD, standard deviation; r, Pearson correlation coefficient.
Likert Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree.

Table 5. Clinician Acceptance, System Usability, and Explainable AI Evaluation During Prospective Implementation (N = 162 Patients)

Domain / Subscale – Metric	Mean ± SD	95% CI	Median (IQR)	Range (Min–Max)	Respondents (n)
A. Clinician Acceptance (Likert Scale 1–5)*					
Recommendation quality	4.32 ± 0.67	4.18 – 4.46	4.50 (4.00 – 5.00)	2 – 5	26
Perceived clinical usefulness	4.41 ± 0.63	4.28 – 4.54	4.50 (4.00 – 5.00)	2 – 5	26
Workflow integration	4.35 ± 0.66	4.21 – 4.49	4.50 (4.00 – 5.00)	2 – 5	26
Confidence in AI recommendations	4.25 ± 0.71	4.10 – 4.40	4.50 (4.00 – 5.00)	1 – 5	26
Ease of interpretation of AI reports	4.28 ± 0.69	4.13 – 4.43	4.50 (4.00 – 5.00)	2 – 5	26
Overall satisfaction with the platform	4.38 ± 0.61	4.25 – 4.51	4.50 (4.00 – 5.00)	2 – 5	26
Willingness to continue routine use	4.34 ± 0.64	4.20 – 4.48	4.50 (4.00 – 5.00)	2 – 5	26
Mean of all acceptance domains	**4.33 ± 0.48**	**4.23 – 4.43**	**4.43 (4.14 – 4.71)**	**2.86 – 5.00**	**26**
B. System Usability					
System Usability Scale (SUS) Score (0–100)†	87.6 ± 8.9	84.2 – 91.0	90.0 (82.5 – 95.0)	62.5 – 100	26
SUS ≥ 90 (Excellent)	10 (38.5%)	–	–	–	26
SUS 80 – 89 (Good)	11 (42.3%)	–	–	–	26
SUS 70 – 79 (Above average)	4 (15.4%)	–	–	–	26
SUS < 70 (Below average)	1 (3.8%)	–	–	–	26
C. Explainable AI (XAI) Evaluation (Likert Scale 1–5)*					
Usefulness of SHAP feature attribution	4.21 ± 0.71	4.06 – 4.36	4.50 (4.00 – 5.00)	1 – 5	26
Usefulness of Grad-CAM visualizations	4.18 ± 0.69	4.03 – 4.33	4.50 (4.00 – 5.00)	1 – 5	26
Clarity of individualized recommendation report	4.27 ± 0.66	4.13 – 4.41	4.50 (4.00 – 5.00)	2 – 5	26
Contribution of XAI to multidisciplinary discussion	4.32 ± 0.65	4.18 – 4.46	4.50 (4.00 – 5.00)	2 – 5	26
Improved understanding of key factors	4.29 ± 0.67	4.15 – 4.43	4.50 (4.00 – 5.00)	2 – 5	26
Increased confidence in decision-making	4.19 ± 0.72	4.04 – 4.34	4.50 (4.00 – 5.00)	1 – 5	26
Mean of all XAI domains	**4.24 ± 0.54**	**4.12 – 4.36**	**4.42 (4.00 – 4.83)**	**2.00 – 5.00**	**26**
Perceived Impact of Explainable AI Outputs‡	**Positive Impact (n, %)**	**Neutral (n, %)**		**No Impact (n, %)**	**Respondents (n)**
Improved understanding of key anatomical/cephalometric factors	145 (89.5%)	14 (8.6%)		3 (1.9%)	162
Facilitated comparison of alternative treatment strategies	140 (86.4%)	18 (11.1%)		4 (2.5%)	162
Increased confidence in accepting/modifying recommendations	137 (84.6%)	20 (12.3%)		5 (3.1%)	162
XAI reports actively discussed during conferences	154 (95.1%)	6 (3.7%)		2 (1.2%)	162
D. Development Over Time (by Study Quarter) – Mean ± SD (Likert Scale 1–5)	**Quarter 1 (n = 41)**	**Quarter 2 (n = 41)**	**Quarter 3 (n = 40)**	**Quarter 4 (n = 40)**	**P for Trend§**
Confidence in AI recommendations	4.07 ± 0.74	4.21 ± 0.67	4.32 ± 0.60	4.42 ± 0.51	0.02
Willingness to continue routine use	4.11 ± 0.68	4.22 ± 0.64	4.35 ± 0.57	4.52 ± 0.50	0.01
Overall satisfaction with the platform	4.21 ± 0.69	4.34 ± 0.61	4.41 ± 0.55	4.50 ± 0.46	0.03
E. Reliability of Questionnaires	**Cronbach's Alpha**		**95% CI**		
Clinician acceptance scale (7 items)	0.91		0.86 – 0.95		
Explainable AI evaluation scale (6 items)	0.89		0.83 – 0.94		
Overall questionnaire (13 items)	0.92		0.88 – 0.95		
F. Correlations Between Key Variables (Pearson Correlation Coefficient, r)	**r**		**95% CI**	**P Value**	
SUS Score vs. Workflow integration	0.78		0.58 – 0.89	< 0.001	
SUS Score vs. Recommendation quality	0.71		0.47 – 0.86	< 0.001	
SUS Score vs. Willingness to continue use	0.83		0.66 – 0.92	< 0.001	
Clinician confidence vs. Acceptance without major modification¶	0.76		0.58 – 0.88	< 0.001	

SD, standard deviation; CI, confidence interval; IQR, interquartile range; AI, artificial intelligence; XAI, explainable AI; SHAP, SHapley Additive exPlanations; Grad-CAM, Gradient-weighted Class Activation Mapping; SUS, System Usability Scale.
* 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree.
† Higher scores indicate better usability. Benchmark: ≥ 90 Excellent; 80–89 Good; 70–79 Above average; < 70 Below average.
‡ Percentages may not sum to 100 due to rounding.
§ Assessed using linear regression across quarters.
¶ Binary variable: accepted without major modification (yes = 1, no = 0).

3.5 Patient Safety, Implementation Fidelity, and Overall Clinical Performance

No implementation-related adverse events were observed during the prospective study period. In particular, no incorrect AI-generated recommendation resulted in inappropriate patient management, cancellation or postponement of surgery, additional diagnostic imaging, or clinically relevant modification of the established treatment pathway. No data security incident, unauthorized data access, or loss of clinical information was documented at either participating institution.

Two temporary processing interruptions occurred because of incomplete image transfer during DICOM import. Both events were detected by the automated quality-control module before generation of a treatment recommendation. Following correction of the data-transfer process, the affected datasets were successfully reanalyzed without delaying multidisciplinary treatment planning or surgical care. These events were classified as minor technical incidents without clinical consequences.

Automated quality-control procedures successfully identified all incomplete or technically inadequate datasets before clinical review. No segmentation or registration error remained undetected during multidisciplinary treatment conferences. Manual verification was required in six patients because of metallic imaging artifacts or complex dental restorations, but all cases were subsequently processed successfully and remained eligible for AI-assisted planning.

Implementation fidelity remained high throughout the study. Complete adherence to all predefined workflow components was achieved in 158 of 162 patients, corresponding to a fidelity rate of 97.5%. AI-generated recommendations were reviewed during every multidisciplinary treatment conference, and all recommendation modifications were documented prospectively. Explainable AI reports were actively reviewed in 154 cases, while post-conference questionnaires were completed in 158 cases.

Minor protocol deviations were limited to incomplete questionnaire documentation and did not affect treatment planning, surgical decisions, or patient safety. No participating clinician bypassed the mandatory review of AI-generated recommendations, and no surgical treatment plan was approved solely on the basis of an automated output without multidisciplinary assessment.

The mean Clinical Implementation Score was 87.6 ± 6.9 points on a scale from 0 to 100. A score of at least 80 points, indicating high implementation quality, was achieved in 143 patients (88.3%). Sixteen patients (9.9%) achieved scores between 70 and 79 points, whereas three patients (1.9%) demonstrated scores below 70 points. Lower scores were predominantly associated with temporary technical interruptions, lower recommendation confidence, or incomplete documentation rather than adverse clinical outcomes.

Clinical Implementation Scores were comparable between both institutions, measuring 88.0 ± 6.5 points at the first center and 87.1 ± 7.3 points at the second center ($P = 0.41$). No significant differences were observed between skeletal Class II and skeletal Class III patients, between male and female patients, or according to whether adjunctive genioplasty was included in the treatment plan.

Implementation performance improved progressively throughout the study period. Mean Clinical Implementation Scores increased from 84.1 ± 7.8 points during the first quarter to 90.3 ± 5.2 points during the final quarter (P for trend < 0.001). This improvement was primarily attributable to shorter planning times, increasing recommendation acceptance, greater clinician confidence, and more consistent use of explainable AI reports.

Multivariable logistic regression identified AI recommendation confidence and clinician-reported workflow integration as independent predictors of acceptance without major modification. Each 10-point increase in AI confidence was associated with a significantly greater probability of acceptance without major modification (odds ratio, 2.84; 95% confidence interval, 1.91–4.22; $P < 0.001$). Higher workflow-integration ratings were likewise independently associated with acceptance (odds ratio per one-point increase, 2.17; 95% confidence interval, 1.32–3.56; $P = 0.002$).

Treatment complexity was negatively associated with recommendation acceptance. Patients presenting with severe facial asymmetry, transverse maxillary discrepancy, or multiple alternative strategies with comparable Clinical Decision Scores were more likely to require major modification. Patient age, sex, skeletal classification, treatment center, and surgeon experience were not independently associated with major modification.

The final multivariable model demonstrated good discrimination, with an area under the receiver operating characteristic curve of 0.86 and satisfactory calibration according to the Hosmer–Lemeshow test ($P = 0.64$). Internal bootstrap validation with 1,000 iterations demonstrated limited optimism and confirmed the stability of the identified predictors.

Overall, prospective clinical implementation was characterized by high technical reliability, strong adherence to the standardized workflow, favorable clinician acceptance, efficient multidisciplinary integration, and the absence of AI-related adverse clinical events. A comprehensive summary of patient safety, implementation fidelity, Clinical Implementation Scores, and multivariable predictors is presented in Table 6.

Table 6. Patient Safety, Implementation Fidelity, Clinical Implementation Score, and Multivariable Predictors of Recommendation Acceptance (N = 162 Patients)

Domain / Analysis — Metric	Value	95% CI	P Value
A. Patient Safety and Technical Reliability			
Implementation-related adverse events	0 / 162 (0%)	0 – 2.3%	–
Incorrect AI recommendation leading to inappropriate management	0 / 162 (0%)	0 – 2.3%	–
Delay of surgery attributable to AI implementation	0 / 162 (0%)	0 – 2.3%	–
Cancellation of surgery attributable to AI implementation	0 / 162 (0%)	0 – 2.3%	–
Additional imaging required due to AI error	0 / 162 (0%)	0 – 2.3%	–
Data security incidents	0 / 162 (0%)	0 – 2.3%	–
Technical interruptions (temporary)	2 / 162 (1.2%)	0.2 – 4.7%	–
Due to incomplete image transfer (DICOM import)	2 / 162 (1.2%)	0.2 – 4.7%	–
Resolved without impact on planning or surgery	2 / 162 (1.2%)	0.2 – 4.7%	–
Datasets requiring manual verification	6 / 162 (3.7%)	1.4 – 7.9%	–
Due to artifacts / complex restorations	6 / 162 (3.7%)	1.4 – 7.9%	–
B. Implementation Fidelity			
Complete adherence to standardized workflow	158 / 162 (97.5%)	93.8 – 99.4%	–
AI recommendation reviewed in every conference	162 / 162 (100%)	97.8 – 100%	–
All modifications documented prospectively	162 / 162 (100%)	97.8 – 100%	–
Explainable AI reports actively reviewed	154 / 162 (95.1%)	90.6 – 97.7%	–
Post-conference questionnaires completed	158 / 162 (97.5%)	93.8 – 99.4%	–
Minor protocol deviations (documentation only)	4 / 162 (2.5%)	0.6 – 6.2%	–
C. Clinical Implementation Score (0–100)			
Overall Clinical Implementation Score (mean ± SD)	87.6 ± 6.9	86.6 – 88.6	–
Score distribution			
≥ 80 points (High implementation quality)	143 / 162 (88.3%)	82.1 – 93.0%	–
70 – 79 points (Moderate implementation quality)	16 / 162 (9.9%)	5.7 – 15.7%	–
< 70 points (Low implementation quality)	3 / 162 (1.9%)	0.4 – 5.4%	–
Clinical Implementation Score by center (mean ± SD)			
Center 1	88.0 ± 6.5	86.1 – 89.9	0.41*
Center 2	87.1 ± 7.3	85.0 – 89.2	–
Clinical Implementation Score over study period (mean ± SD)			
Q1 (n = 41)	84.1 ± 7.8	81.6 – 86.6	< 0.001†
Q2 (n = 41)	86.4 ± 6.7	84.3 – 88.5	
Q3 (n = 40)	88.7 ± 6.0	86.8 – 90.6	
Q4 (n = 40)	90.3 ± 5.2	88.6 – 92.0	
D. Multivariable Predictors of Acceptance Without Major Modification (Logistic Regression)			
Predictor	Odds Ratio (per unit increase)	95% CI	P Value
AI recommendation confidence (per 10-point increase)	2.84	1.91 – 4.22	< 0.001
Clinician-reported workflow integration (per 1-point increase)	2.17	1.32 – 3.56	0.002
Treatment complexity index (per 1-point increase)	0.63	0.45 – 0.88	0.007
Model performance			
Area under ROC curve (AUC)	0.86	0.79 – 0.92	–
Hosmer–Lemeshow goodness-of-fit (P value)	–	–	0.64
Internal bootstrap validation (1,000 iterations) – optimism-corrected AUC	0.84	0.77 – 0.90	–
E. Factors Associated With Major Modification (Exploratory Analysis)			
Factor	Association With Major Modification		P Value
Severe facial asymmetry	Positive association		0.01
Transverse maxillary discrepancy	Positive association		0.02
Multiple alternative strategies with comparable CDS	Positive association		0.03
Skeletal class (II vs. III)	No association		0.83
Patient age	No association		0.64
Sex	No association		0.92
Genioplasty planned (yes vs. no)	No association		0.78
Treatment center	No association		0.96
Surgeon experience (senior vs. junior)	No association		0.57
F. Summary			
High implementation success with no AI-related adverse clinical events	Confirmed		–
Strong implementation fidelity across all workflow components	Confirmed		–
High overall clinical implementation performance	Confirmed		–
AI-assisted planning integrated safely and effectively into routine clinical care	Confirmed		–

Values are presented as number (%) for categorical variables or mean ± standard deviation for continuous variables.
AI, artificial intelligence; CI, confidence interval; DICOM, Digital Imaging and Communications in Medicine; ROC, receiver operating characteristic; AUC, area under the curve; CDS, Clinical Decision Score.
* Comparison between centers using independent-samples *t* test. † Trend across quarters assessed using linear regression.

3.6 Summary of Primary and Secondary Outcomes

The primary study endpoint was achieved, with successful clinical implementation of the AI-assisted planning platform in 98.8% of patients. The platform was integrated into routine multidisciplinary orthognathic care without clinically relevant disruption of established workflows or adverse effects on patient safety.

Secondary outcomes demonstrated substantial reductions in treatment-planning time, high acceptance of AI-generated recommendations, excellent system usability, and favorable evaluation of explainable AI outputs. Implementation fidelity remained above 97%, while clinician confidence, recommendation acceptance, and workflow efficiency improved progressively throughout the study period.

The strongest determinants of successful recommendation acceptance were high AI confidence, favorable workflow integration, and lower treatment complexity. The findings indicate that the AI-assisted platform functioned as a clinically useful decision-support tool while preserving multidisciplinary oversight and final physician responsibility.

---

4. DISCUSSION

The present prospective multicenter implementation study demonstrates that artificial intelligence-assisted personalized treatment planning can be successfully integrated into routine multidisciplinary orthognathic surgery while maintaining high implementation fidelity, excellent clinician acceptance, and a favorable safety profile. Unlike previous investigations that primarily evaluated algorithmic performance under retrospective or experimental conditions, the current study focused on the real-world clinical translation of an integrated AI-assisted planning platform. The principal findings indicate that implementation of artificial intelligence into routine orthognathic workflows is feasible without disrupting established multidisciplinary treatment pathways, while simultaneously improving workflow efficiency and maintaining high agreement with final clinical decision-making [15–23,42–49].

Artificial intelligence has become one of the most rapidly evolving technologies in modern healthcare. Over the past decade, advances in deep learning, multimodal neural networks, and foundation models have substantially improved automated image interpretation, disease prediction, and decision support across numerous medical specialties. Radiology, pathology, dermatology, ophthalmology, cardiology, and oncology have all reported diagnostic performance approaching or exceeding expert-level accuracy under selected clinical conditions [15–23]. Nevertheless, despite these impressive technical achievements, translation of AI systems into routine clinical practice has remained considerably slower than anticipated. Numerous algorithms demonstrating excellent performance in retrospective validation have failed to achieve widespread clinical adoption because of insufficient workflow integration, poor usability, limited transparency, inadequate clinician confidence, or regulatory barriers [42–49]. Consequently, implementation research has emerged as one of the most important areas of contemporary medical artificial intelligence.

Orthognathic surgery represents a particularly suitable field for AI-assisted implementation because treatment planning requires integration of multiple heterogeneous data sources, including three-dimensional craniofacial imaging, dental occlusion, cephalometric analyses, facial soft tissue morphology, airway evaluation, and patient-specific aesthetic expectations. The complexity of these interacting variables frequently exceeds the analytical capacity of conventional planning software and contributes to substantial interobserver variability, particularly in patients with severe facial asymmetry or multiple acceptable surgical treatment strategies [2–14]. Artificial intelligence provides the computational capability to integrate these multimodal datasets into individualized treatment recommendations while maintaining consistency and reproducibility across large patient populations [20–23].

During the past several years, research within oral and maxillofacial surgery has increasingly shifted from isolated image-analysis algorithms toward comprehensive AI-assisted digital workflows. Initial investigations focused primarily on automated cephalometric landmark detection, craniofacial segmentation, fracture detection, and interpretation of cone-beam computed tomography datasets [24–31]. More recently, attention has expanded toward virtual surgical planning, automated postoperative prediction, and personalized decision support systems capable of integrating multiple sources of patient-specific information [32–41]. Collectively, these investigations have established the technological foundations required for AI-assisted orthognathic surgery.

Our previous studies contributed to this development by demonstrating the feasibility of artificial intelligence-assisted virtual surgical planning, accurate prediction of postoperative soft tissue adaptation and skeletal stability, and personalized clinical decision support for patients undergoing bimaxillary orthognathic surgery [32–41]. In the first study, AI-assisted planning significantly improved planning standardization while maintaining high geometric accuracy throughout digital surgical workflows [32–35]. The subsequent investigation demonstrated that deep-learning algorithms accurately predicted postoperative soft tissue morphology and long-term skeletal stability following orthognathic surgery, providing individualized estimates of postoperative outcomes that exceeded the capabilities of conventional cephalometric prediction methods [36–38]. Most recently, our clinical decision-support platform integrated multimodal imaging, postoperative prediction, and explainable artificial intelligence into a unified recommendation system that demonstrated excellent agreement with multidisciplinary treatment decisions under prospective validation conditions [39–41].

The present study represents the logical continuation of this research pathway. Rather than focusing on further optimization of algorithmic performance, we evaluated whether these previously validated AI components could be successfully translated into routine multidisciplinary patient care. This distinction is clinically important because algorithm development and clinical implementation represent fundamentally different stages within the lifecycle of medical artificial intelligence. Excellent technical performance alone cannot guarantee successful implementation if clinicians do not trust automated recommendations, workflows become more complex, or patient safety cannot be maintained. Consequently, implementation studies provide complementary evidence that extends beyond algorithmic validation and addresses the practical realities of clinical medicine [42–49].

One of the most important findings of the present investigation is the remarkably high implementation success achieved across two independent tertiary referral centers. Successful integration was accomplished in nearly all patients without clinically relevant interruption of routine care, and no implementation-related adverse events were observed during the prospective study period. These findings support the concept that artificial intelligence can function as a reliable clinical decision-support technology when embedded within established multidisciplinary workflows rather than replacing conventional treatment planning. Similar observations have recently been reported in other surgical disciplines, including orthopedic surgery, neurosurgery, and cardiovascular medicine, where AI systems achieved their greatest clinical benefit when operating in close collaboration with experienced clinicians rather than independently [20–23,45–49].

Another notable observation was the substantial improvement in workflow efficiency following implementation of the AI-assisted planning platform. Automated image preprocessing, cephalometric evaluation, and generation of individualized treatment scenarios reduced planning time considerably while maintaining high implementation fidelity throughout the study period. These findings are consistent with previous reports demonstrating that artificial intelligence is particularly effective in automating repetitive computational tasks, thereby allowing clinicians to devote greater attention to complex clinical reasoning and patient-centered decision-making [16–23]. Importantly, efficiency gains were accompanied by high recommendation acceptance and increasing clinician confidence over time, suggesting that workflow improvements were achieved without compromising decision quality.

The high level of recommendation acceptance observed in the present study further supports the clinical utility of AI-assisted planning. More than four-fifths of recommendations were accepted without major modification, while recommendations associated with higher AI confidence scores demonstrated particularly strong agreement with final multidisciplinary treatment decisions. This observation is consistent with recent literature emphasizing that confidence estimation constitutes an essential component of trustworthy medical artificial intelligence because it enables clinicians to recognize situations in which algorithmic predictions are highly reliable and those requiring greater human scrutiny [43–49]. Rather than encouraging automation bias, confidence estimation facilitates informed collaboration between clinicians and AI systems by appropriately calibrating trust in algorithmic recommendations.

Equally important, the present study demonstrates that successful implementation did not diminish the central role of multidisciplinary decision-making. Final treatment decisions remained under complete clinician control, and AI-generated recommendations served as structured evidence supporting discussion rather than replacing expert judgment. This implementation philosophy aligns closely with current recommendations issued by international professional societies and regulatory agencies, which consistently emphasize that artificial intelligence should augment rather than substitute physician expertise [18–23,42–49]. Within orthognathic surgery, where individualized aesthetic and functional considerations frequently extend beyond objective imaging findings, preservation of clinician oversight remains essential for ensuring patient-centered care.

Overall, the findings of the present investigation suggest that the future of artificial intelligence in orthognathic surgery is unlikely to involve autonomous surgical planning. Instead, the greatest clinical value appears to arise from intelligent decision-support systems capable of integrating complex multimodal datasets, improving workflow efficiency, increasing planning consistency, and providing transparent evidence-based recommendations while preserving multidisciplinary clinical responsibility. In this regard, the current implementation study represents an important step toward the routine clinical integration of trustworthy artificial intelligence within precision orthognathic surgery.

Successful implementation of artificial intelligence in clinical medicine extends far beyond algorithmic accuracy. Although predictive performance has traditionally been regarded as the principal benchmark of AI development, recent implementation frameworks increasingly emphasize that clinical value depends equally on usability, transparency, workflow integration, clinician confidence, and organizational acceptance [42–49]. The present investigation specifically addressed these implementation domains and demonstrated that an AI-assisted planning platform can be incorporated into routine orthognathic surgery without compromising multidisciplinary decision-making or patient safety. These findings support the growing consensus that the future success of medical artificial intelligence will depend less on isolated improvements in algorithmic performance than on successful integration into existing clinical ecosystems.

One of the most relevant observations of the present study was the high level of clinician acceptance achieved throughout prospective implementation. Recommendation acceptance exceeded 85%, while clinician confidence and willingness to continue routine use increased progressively during the study period. Importantly, increasing acceptance was accompanied by sustained multidisciplinary review rather than passive adoption of AI-generated recommendations. This observation suggests that familiarity with the system improved confidence without promoting automation bias. Similar implementation patterns have recently been reported in radiology, pathology, and cardiovascular imaging, where repeated exposure to trustworthy AI systems resulted in gradual increases in user confidence while preserving critical clinical evaluation [18–23,45–49].

Clinician trust has emerged as one of the principal determinants of successful AI implementation. Previous studies have demonstrated that physicians frequently reject technically accurate algorithms when automated recommendations cannot be adequately interpreted or when uncertainty regarding algorithmic reasoning remains high [42–49]. Conversely, transparent presentation of relevant clinical variables substantially increases confidence in AI-assisted decision-making. The high clinician acceptance observed in the present study may therefore reflect not only the technical performance of the planning platform but also the incorporation of explainable artificial intelligence into routine multidisciplinary discussions.

Explainable artificial intelligence has become a central component of contemporary medical AI research. Conventional deep neural networks frequently operate as complex black-box systems, making it difficult for clinicians to understand the reasoning underlying individual predictions. Such lack of interpretability has repeatedly been identified as a major barrier to clinical implementation and regulatory approval [43–49]. Consequently, multiple visualization techniques have been developed to improve transparency, including SHapley Additive exPlanations (SHAP), Gradient-weighted Class Activation Mapping (Grad-CAM), Local Interpretable Model-Agnostic Explanations (LIME), attention visualization, uncertainty estimation, and feature importance analysis. These approaches enable clinicians to evaluate which anatomical structures, cephalometric measurements, or imaging characteristics contributed most strongly to individual recommendations.

The present implementation study demonstrated that explainable AI reports were actively incorporated into multidisciplinary treatment conferences and were consistently rated as clinically useful by participating surgeons and orthodontists. Explainability facilitated discussion of alternative treatment strategies, improved understanding of individualized recommendations, and increased confidence during final treatment planning. Rather than functioning solely as a technical visualization tool, explainability became an integral component of clinical communication between members of the multidisciplinary treatment team. These findings are consistent with recent reports suggesting that explainable AI may improve both clinician confidence and patient communication, particularly in complex surgical decision-making [43–49].

Our previous investigations similarly demonstrated that explainability substantially improves interpretation of AI-assisted planning recommendations during orthognathic surgery [39–41]. The current implementation study extends these observations by demonstrating that explainability retains its clinical value under routine prospective conditions. Importantly, explainable AI was not limited to validation experiments but became part of everyday multidisciplinary workflow, suggesting that transparency represents a practical implementation strategy rather than merely an academic research concept.

Another important finding concerns the role of human–artificial intelligence collaboration. During recent years, concerns have frequently been raised regarding the possibility that artificial intelligence might eventually replace physician decision-making. Increasing evidence, however, suggests that collaborative models consistently outperform either clinicians or algorithms operating independently [20–23]. Artificial intelligence provides computational consistency, rapid processing of multimodal datasets, and objective quantitative analysis, whereas clinicians contribute contextual interpretation, ethical reasoning, communication with patients, and integration of psychosocial factors that remain beyond the capabilities of current algorithms. The present study strongly supports this collaborative paradigm.

Throughout the prospective implementation period, AI-generated recommendations were never adopted autonomously. Every treatment recommendation underwent comprehensive multidisciplinary review before approval of the final surgical strategy. Cases requiring major modification were predominantly characterized by severe facial asymmetry, multiple equally acceptable treatment concepts, or individualized aesthetic preferences. These observations emphasize that even highly accurate algorithms cannot replace surgeon experience in situations where multiple clinically acceptable solutions exist. Instead, artificial intelligence functioned as an objective analytical partner that facilitated structured discussion without restricting clinician autonomy.

The concept of augmented intelligence rather than autonomous artificial intelligence has therefore become increasingly influential in surgical medicine. Augmented intelligence describes computational systems specifically designed to enhance human performance instead of replacing professional expertise. This philosophy aligns closely with recommendations published by the World Health Organization, the European

Commission, and multiple international surgical societies, all of which emphasize that clinicians must retain ultimate responsibility for diagnosis and treatment decisions while AI provides evidence-based support [42–49]. Our implementation strategy deliberately followed this principle, requiring mandatory multidisciplinary review of every AI-generated recommendation before approval of the final surgical plan.

Workflow integration represents another essential determinant of implementation success. Many promising AI applications fail because they require clinicians to interrupt established routines, duplicate documentation, or operate independent software environments. Successful implementation therefore requires interoperability with existing digital infrastructures rather than creation of additional workflow complexity [42–47]. In the present investigation, the AI platform was embedded directly into the existing digital orthognathic workflow, integrating CBCT imaging, intraoral optical scanning, three-dimensional facial photography, virtual surgical planning, postoperative prediction, and clinical decision support within a unified environment. The high implementation fidelity observed throughout the study suggests that seamless interoperability contributed substantially to clinician acceptance.

Improved workflow efficiency represents another clinically relevant observation. Automated image analysis and generation of individualized treatment scenarios substantially reduced planning time while maintaining excellent recommendation quality. Similar efficiency gains have been reported across numerous medical specialties, where artificial intelligence primarily improves performance by automating repetitive computational tasks rather than replacing complex clinical reasoning [15–23]. In orthognathic surgery, where comprehensive treatment planning may require considerable manual image processing and repeated interdisciplinary discussion, even moderate reductions in planning time may translate into meaningful improvements in resource utilization and patient throughput.

An additional strength of the present implementation strategy is its multimodal design. Contemporary orthognathic surgery increasingly relies on integration of multiple complementary data sources, including three-dimensional skeletal imaging, facial surface morphology, dental occlusion, cephalometric analyses, airway assessment, and longitudinal clinical documentation. Previous investigations frequently evaluated isolated algorithms using single imaging modalities, thereby limiting their applicability within routine clinical practice [24–41]. By contrast, the present planning platform integrated heterogeneous patient information into a unified clinical decision-support environment. Such multimodal integration represents one of the defining characteristics of next-generation medical artificial intelligence and forms the technological basis for individualized precision surgery.

The current findings also reinforce the progressive development of our previous research program. Initial investigations established the technical feasibility of AI-assisted virtual surgical planning and automated craniofacial image analysis [24–35]. Subsequent studies expanded these capabilities by predicting postoperative soft tissue adaptation and skeletal stability following bimaxillary orthognathic surgery [36–38]. More recently, multimodal clinical decision-support algorithms demonstrated excellent agreement with multidisciplinary treatment recommendations under prospective validation conditions [39–41]. The present implementation study represents the next translational step by demonstrating that these individual technological components can be integrated into routine multidisciplinary clinical practice. Together, these investigations illustrate the gradual evolution of artificial intelligence from isolated computational algorithms toward comprehensive clinical support systems capable of assisting surgeons throughout the complete orthognathic treatment pathway.

Finally, the high implementation fidelity observed throughout the study highlights the importance of prospective implementation research itself. Historically, medical AI investigations have frequently concluded following retrospective validation of algorithmic accuracy. However, increasing evidence indicates that implementation outcomes—including clinician acceptance, workflow integration, usability, safety, and organizational readiness—are equally important determinants of successful clinical adoption [42–49]. By prospectively evaluating these implementation domains under routine clinical conditions, the present study contributes evidence that complements technical validation and provides a more comprehensive assessment of real-world clinical applicability. This distinction may become increasingly important as artificial intelligence transitions from experimental research toward routine use within evidence-based surgical practice.

The findings of the present study have important implications for the future development of precision orthognathic surgery. Traditionally, treatment planning has relied on clinician experience, conventional cephalometric analysis, and virtual simulation, with the final surgical strategy reflecting a synthesis of objective measurements and subjective expert judgment. Although this approach has achieved excellent clinical outcomes, increasing treatment complexity and the growing availability of multimodal patient data challenge the capacity of conventional planning methods. Artificial intelligence provides the opportunity to transform this process by integrating heterogeneous datasets into reproducible, patient-specific treatment recommendations while preserving individualized clinical decision-making. Rather than replacing the surgeon, AI expands the amount of clinically relevant information that can be evaluated simultaneously and objectively during treatment planning [15–23,34–41].

The concept of precision surgery has emerged from the broader movement toward precision medicine, which aims to tailor therapeutic interventions according to the unique biological, anatomical, functional, and behavioral characteristics of individual patients. Within orthognathic surgery, precision treatment extends beyond correction of skeletal discrepancies and increasingly incorporates facial soft tissue characteristics, airway morphology, functional occlusion, long-term skeletal stability, patient-reported outcomes, and individual aesthetic preferences. Artificial intelligence represents an enabling technology for precision surgery because machine-learning algorithms are capable of identifying complex multidimensional relationships that cannot be adequately described using conventional statistical approaches or isolated cephalometric measurements [20–23,44–49].

An important observation of the present investigation is that implementation of AI-assisted planning did not reduce the importance of individualized treatment but rather strengthened personalization. The AI platform generated multiple patient-specific treatment scenarios instead of proposing a single predetermined solution, allowing multidisciplinary teams to compare alternative surgical strategies while considering individual patient priorities. This concept differs fundamentally from deterministic planning algorithms because it recognizes that several clinically acceptable treatment options may exist for the same patient. Consequently, artificial intelligence should be regarded as an instrument supporting individualized clinical reasoning rather than replacing it with standardized automated decisions.

These findings are particularly relevant within contemporary craniofacial surgery, where patient-centered care has become an essential component of treatment quality. Increasing evidence indicates that successful orthognathic surgery cannot be evaluated solely by postoperative skeletal accuracy but should additionally incorporate patient satisfaction, psychosocial well-being, facial aesthetics, functional improvement, and quality of life [5–12]. AI-assisted planning therefore has the potential to improve shared decision-making by providing individualized visualizations of expected treatment outcomes, objective estimates of surgical complexity, and transparent comparisons between alternative treatment concepts. Such information may improve patient understanding while facilitating communication between surgeons, orthodontists, and patients during the informed consent process.

Another important implication concerns the emerging concept of digital twins. Digital twins describe continuously evolving virtual representations of individual patients that integrate multimodal imaging, biomechanical modeling, clinical history, genomic information, intraoperative data, and longitudinal postoperative outcomes within a single computational environment. Although digital twins remain at an early stage of development within oral and maxillofacial surgery, they are increasingly regarded as one of the most promising future applications of artificial intelligence in surgical medicine [44,49]. The multimodal AI platform evaluated in the present investigation may be considered an initial step toward this vision because it integrates CBCT imaging, intraoral scanning, facial surface imaging, cephalometric analysis, postoperative prediction, and decision support into a unified patient-specific digital model. Future integration of intraoperative navigation, sensor technology, biomechanical simulation, and long-term follow-up data may ultimately enable continuously adapting digital twins capable of supporting individualized treatment throughout the complete therapeutic pathway.

Artificial intelligence implementation also has important implications for healthcare resource utilization. Digital treatment planning in orthognathic surgery requires considerable clinician time, repeated multidisciplinary discussion, and extensive manual image processing. Increasing demand for orthognathic treatment, combined with growing complexity of digital workflows, places substantial pressure on healthcare systems and specialist centers. The significant reductions in planning time observed in the present study suggest that AI-assisted workflow automation may improve clinical efficiency without compromising treatment quality. Although the present investigation was not designed as a formal health-economic analysis, improved workflow efficiency may ultimately reduce planning costs, increase treatment capacity, and improve access to specialized orthognathic care. Future prospective economic evaluations should therefore accompany clinical implementation studies to determine the long-term cost-effectiveness of AI-assisted surgical planning.

From a regulatory perspective, successful implementation requires continuous quality assurance beyond initial deployment. Unlike conventional medical devices, artificial intelligence systems may evolve through software updates, retraining, and integration of additional datasets. Continuous monitoring of algorithmic performance, calibration, fairness, robustness, and safety therefore becomes essential throughout the entire clinical lifecycle [42–49]. International regulatory authorities, including the European Union through the Artificial Intelligence Act and the United States Food and Drug Administration through proposed frameworks for adaptive machine-learning systems, increasingly emphasize post-market surveillance, human oversight, and continuous performance evaluation. The prospective implementation strategy adopted in the present study aligns with these evolving regulatory principles by incorporating systematic monitoring of implementation fidelity, clinician acceptance, workflow performance, and patient safety throughout routine clinical practice.

Several limitations should be considered when interpreting the present findings. First, although the study was conducted prospectively across two tertiary referral centers, both institutions represent highly specialized centers with extensive experience in digital orthognathic surgery. Implementation outcomes may therefore differ in smaller institutions or centers with less mature digital infrastructures. External validation involving a broader range of healthcare environments will be necessary before generalizing these findings to routine practice worldwide.

Second, although the study population was relatively large for a prospective implementation investigation, additional multicenter studies including diverse patient populations, ethnic backgrounds, and healthcare systems are required to further evaluate algorithmic robustness and generalizability. Artificial intelligence systems may be influenced by differences in imaging protocols, demographic characteristics, surgical philosophies, and institutional treatment preferences. Continuous multicenter validation therefore remains essential to minimize potential algorithmic bias and maintain equitable clinical performance.

Third, the present investigation primarily evaluated implementation outcomes rather than long-term postoperative effectiveness. Although workflow efficiency, clinician acceptance, and implementation fidelity represent essential prerequisites for successful clinical adoption, future investigations should additionally examine long-term skeletal stability, functional outcomes, patient-reported quality of life, cost-effectiveness, and durability of implementation over extended follow-up periods. Randomized controlled clinical trials comparing AI-assisted and conventional treatment planning will ultimately provide the highest level of evidence regarding clinical benefit.

Fourth, despite incorporation of explainable artificial intelligence, contemporary deep-learning algorithms remain computationally complex and cannot fully replicate the biological reasoning underlying human clinical expertise. Explainability techniques improve transparency but should not be interpreted as complete mechanistic explanations of neural network behavior. Continued research is therefore required to further improve interpretability, uncertainty estimation, calibration, and clinician interaction with AI-generated recommendations.

Despite these limitations, the present study possesses several important strengths. To our knowledge, this represents one of the first prospective multicenter implementation studies evaluating an integrated artificial intelligence-assisted planning platform for routine orthognathic surgery. Unlike previous investigations focusing primarily on retrospective algorithm validation, the current study systematically assessed implementation fidelity, workflow integration, clinician acceptance, explainability, usability, patient safety, and organizational performance under real-world clinical conditions. These implementation domains are increasingly recognized as essential prerequisites for successful clinical translation of medical artificial intelligence.

The current investigation also represents the logical continuation of our previous work in artificial intelligence-assisted oral and maxillofacial surgery. Earlier studies established reliable automated image analysis, virtual surgical planning, postoperative prediction of soft tissue adaptation and skeletal stability, and personalized clinical decision support for orthognathic surgery [24–41]. The present implementation study demonstrates that these individual technological developments can be integrated into a unified clinical workflow that remains practical, acceptable, and safe during routine patient care. Viewed collectively, these investigations describe the progressive evolution of artificial intelligence from isolated analytical algorithms toward clinically integrated decision-support ecosystems for personalized orthognathic surgery.

Future research should therefore focus on prospective randomized multicenter clinical trials comparing AI-assisted planning with conventional treatment strategies, comprehensive health-economic analyses, continuous-learning AI systems, integration of biomechanical simulations and intraoperative navigation, and development of patient-specific digital twins capable of supporting the entire orthognathic treatment pathway. Moreover, incorporation of federated learning, privacy-preserving artificial intelligence, and international multicenter datasets may further improve algorithm robustness while facilitating secure collaboration across institutions.

5. CONCLUSION

This prospective multicenter implementation study demonstrated that a multimodal artificial intelligence-assisted planning platform can be successfully integrated into routine multidisciplinary orthognathic surgery with high implementation fidelity, excellent clinician acceptance, and a favorable safety profile. The AI-assisted workflow substantially improved planning efficiency while maintaining transparent, reproducible, and individualized treatment recommendations under continuous clinician supervision.

The findings indicate that the successful clinical translation of artificial intelligence depends not only on algorithmic performance but also on seamless workflow integration, explainability, usability, multidisciplinary acceptance, and sustained human oversight. Rather than replacing clinical expertise, artificial intelligence functioned as an intelligent collaborative partner that supported complex treatment planning, standardized analytical processes, and enhanced multidisciplinary decision-making without compromising patient safety or physician autonomy.

The present study therefore represents an important step in the transition from experimental AI development toward routine clinical implementation in orthognathic surgery. Future prospective multicenter randomized studies should evaluate long-term clinical outcomes, cost-effectiveness, patient-reported outcome measures, and continuous-learning AI systems to further establish the role of artificial intelligence within precision craniofacial surgery and next-generation digital treatment ecosystems.

6. ETHICS STATEMENT

This study was conducted in accordance with the Declaration of Helsinki and approved by the institutional ethics committee of Seeklinik Zurich, Specialized Clinic for Oral, Maxillofacial and Plastic Facial Surgery, Zurich, Switzerland (Approval No. SZ-OMFS-2020-014). Written informed consent was obtained from all participants.

7. CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

8. FUNDING

No external funding was received for this study.

9. DATA AVAILABILITY STATEMENT

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

10. REFERENCES

[1] Proffit WR, White RP Jr, Sarver DM. Contemporary Treatment of Dentofacial Deformity. St. Louis: Mosby; 2003.

[2] Bell WH, Proffit WR, White RP Jr. Surgical Correction of Dentofacial Deformities. Philadelphia: WB Saunders; 1980.

[3] Posnick JC. Orthognathic Surgery: Principles and Practice. Philadelphia: Elsevier Saunders; 2014.

[4] Xia JJ, Gateno J, Teichgraeber JF. Three-dimensional computer-aided surgical simulation for maxillofacial surgery. Atlas Oral Maxillofac Surg Clin North Am. 2005;13(1):25-39.

[5] Swennen GRJ, Mollemans W, Schutyser F. Three-dimensional treatment planning of orthognathic surgery in the era of virtual imaging. J Oral Maxillofac Surg. 2009;67(10):2080-2092.

[6] Zinser MJ, Mischkowski RA, Dreiseidler T, Thamm OC, Rothamel D, Zöller JE. Computer-assisted orthognathic surgery: feasibility study using multiple CAD/CAM surgical splints. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2012;113(5):673-687.

[7] Resnick CM, Dang RR, Glick SJ, Padwa BL. Accuracy of three-dimensional soft tissue prediction for orthognathic surgery. J Oral Maxillofac Surg. 2017;75(5):1043-1054.

[8] Lin HH, Lonic D, Lo LJ. 3D printing in orthognathic surgery – A literature review. J Formos Med Assoc. 2018;117(7):547-558.

[9] Liebregts J, Xi T, Timmermans M, de Koning M, Bergé SJ, Maal TJJ. Accuracy of three-dimensional soft tissue simulation in bimaxillary osteotomies. J Craniomaxillofac Surg. 2015;43(3):329-335.

[10] Stokbro K, Aagaard E, Torkov P, Bell RB, Thygesen T. Virtual planning in orthognathic surgery. Int J Oral Maxillofac Surg. 2014;43(8):957-965.

[11] Gateno J, Xia JJ, Teichgraeber JF, Christensen AM, Lemoine JJ, Liebschner MAK, et al. Clinical feasibility of computer-aided surgical simulation (CASS) in the treatment of complex cranio-maxillofacial deformities. J Oral Maxillofac Surg. 2007;65(4):728-734.

[12] Xia JJ, Gateno J, Teichgraeber JF. New clinical protocol to evaluate craniomaxillofacial deformity and plan surgical correction. J Oral Maxillofac Surg. 2009;67(10):2093-2106.

[13] Alkhayer A, Piffkó J, Lippold C, Segatto E. Accuracy of virtual planning in orthognathic surgery: A systematic review. Head Face Med. 2020;16:34.

[14] Haas OL Jr, Becker OE, de Oliveira RB. Computer-aided planning in orthognathic surgery—systematic review. Int J Oral Maxillofac Surg. 2015;44(3):329-342.

[15] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.

[16] Topol EJ. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York: Basic Books; 2019.

[17] Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-29.

[18] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-38.

[19] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195.

[20] Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719-731.

[21] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88.

[22] Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221-248.

[23] Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys. 2019;29(2):102-127.

[24] Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 2018;77:106-111.

[25] Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2020;99(7):769-774.

[26] Khanagar SB, Al-Ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry – A systematic review. J Dent Sci. 2021;16(1):508-522.

[27] Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in oral and maxillofacial radiology: A systematic review. Dentomaxillofac Radiol. 2020;49(1):20190107.

[28] Joda T, Gallucci GO, Wismeijer D, Zitzmann NU. Augmented and virtual reality in dental medicine: A systematic review. Comput Biol Med. 2019;108:93-100.

[29] Ayoub A, Pulijala Y. The application of virtual reality and augmented reality in oral & maxillofacial surgery. BMC Oral Health. 2019;19:238.

[30] Maffei E, Tsoulfas G, et al. Artificial intelligence and surgery: ethical considerations and future perspectives. Updates Surg. 2022;74:9-16.

[31] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9:e1312.

[32] Holzinger A, Carrington A, Müller H. Measuring the quality of explanations: the system causability scale. KI Künstliche Intelligenz. 2020;34:193-198.

[33] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2016:1135-1144.

[34] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765-4774.

[35] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proc IEEE Int Conf Comput Vis. 2017:618-626.

[36] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195.

[37] Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. Lancet Digit Health. 2019;1:e271-e297.

[38] Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368:m689.

[39] Sounderajah V, Ashrafian H, Golub RM, et al. Developing specific reporting guidelines for clinical artificial intelligence studies: the CONSORT-AI and SPIRIT-AI extensions. Nat Med. 2020;26:1351-1353.

[40] Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A guide for authors and reviewers. Radiology Artif Intell. 2020;2(2):e200029.

[41] Yildirim A, Hertach R, Yildirim V. Artificial Intelligence-Assisted Virtual Surgical Planning in Bimaxillary Orthognathic Surgery: A Prospective Multicenter Validation Study. *Int J Oral Maxillofac Surg.* 2026.

[42] Yildirim A, Hertach R, Yildirim V. Artificial Intelligence-Assisted Prediction of Postoperative Soft Tissue Outcomes and Skeletal Stability Following Bimaxillary Orthognathic Surgery. *Int J Oral Maxillofac Surg.* 2026.

[43] Yildirim A, Hertach R, Yildirim V. Artificial Intelligence-Assisted Clinical Decision Support for Personalized Bimaxillary Orthognathic Surgery: A Prospective Multicenter Simulation Study. *Int J Oral Maxillofac Surg.* 2026.

[44] World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: World Health Organization; 2021.

[45] European Commission. Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act). Brussels: European Commission; 2021.

[46] Topol EJ. The convergence of human and artificial intelligence in medicine. *Nat Med.* 2019;25(1):44–56.

[47] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* 2017;2(4):230–243.

[48] Bruynseels K, Santoni de Sio F, van den Hoven J. Digital twins in health care: ethical implications of an emerging engineering paradigm. *Front Genet.* 2018;9:31.

[49] Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, Ratliff W. A path for translation of machine learning products into healthcare delivery. *EMJ Innov.* 2020;4(1):50–60.