

Article

Early Oral Cancer Detection with AI: Design and Implementation of a Deep Learning Image-Based Chatbot

Pablo Ormeño-Arriagada ^{1,*} , Gastón Márquez ² , Carla Taramasco ³ , Gustavo Gatica ⁴ ,
Juan Pablo Vasconez ⁴  and Eduardo Navarro ⁴ 

¹ Facultad de Ingeniería, Negocios y Ciencias AgroAmbientales, Universidad de Viña del Mar, Viña del Mar 2580022, Chile

² Departamento de Ciencias de la Computación y Tecnología de la Información, Universidad del Bío-Bío, Chillán 4081112, Chile; gmarquez@ubiobio.cl

³ Centro para la Prevención y el Control del Cáncer (CECAN), Facultad de Ingeniería, Universidad Andrés Bello, Viña del Mar 2520000, Chile; carla.taramasco@unab.cl

⁴ Facultad de Ingeniería, Universidad Andrés Bello, Viña del Mar 2520000, Chile; ggatica@unab.cl (G.G.); juan.vasconez@unab.cl (J.P.V.); enavarro@unab.cl (E.N.)

* Correspondence: pablo.ormeno@uvm.cl

Abstract

Oral cancer remains a critical global health challenge, with delayed diagnosis driving high morbidity and mortality. Despite progress in artificial intelligence, computer vision, and medical imaging, early detection tools that are accessible, explainable, and designed for patient engagement remain limited. This study presents a novel system that combines a patient-centred chatbot with a deep learning framework to support early diagnosis, symptom triage, and health education. The system integrates convolutional neural networks, class activation mapping, and natural language processing within a conversational interface. Five deep learning models were evaluated (CNN, DenseNet121, DenseNet169, DenseNet201, and InceptionV3) using two balanced public datasets. Model performance was assessed using accuracy, sensitivity, specificity, diagnostic odds ratio (DOR), and Cohen's Kappa. InceptionV3 consistently outperformed the other models across these metrics, achieving the highest diagnostic accuracy (77.6%) and DOR (20.67), and was selected as the core engine of the chatbot's diagnostic module. The deployed chatbot provides real-time image assessments and personalised conversational support via multilingual web and mobile platforms. By combining automated image interpretation with interactive guidance, the system promotes timely consultation and informed decision-making. It offers a prototype for a chatbot, which is scalable and serves as a low-cost solution for underserved populations and demonstrates strong potential for integration into digital health pathways. Importantly, the system is not intended to function as a formal screening tool or replace clinical diagnosis; rather, it provides preliminary guidance to encourage early medical consultation and informed health decisions.

Keywords: artificial intelligence; chatbot; deep learning; healthcare technology; natural language processing; oral cancer; patient support



Received: 28 July 2025

Revised: 30 September 2025

Accepted: 30 September 2025

Published: 7 October 2025

Citation: Ormeño-Arriagada, P.; Márquez, G.; Taramasco, C.; Gatica, G.; Vasconez, J.P.; Navarro, E. Early Oral Cancer Detection with AI: Design and Implementation of a Deep Learning Image-Based Chatbot. *Appl. Sci.* **2025**, *15*, 10792. <https://doi.org/10.3390/app151910792>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tools that support early disease detection through patient interaction have become increasingly important in modern healthcare. When enhanced by artificial intelligence (AI), these tools provide rapid, cost-effective, and accessible solutions for identifying potentially

life-threatening conditions. Chatbots, in particular, present a valuable means of enabling early detection due to their broad accessibility, with the potential to prevent thousands of deaths annually [1]. The integration of AI and deep learning (DL) technologies further strengthens these capabilities, as both have demonstrated strong performance in analysing quantitative and qualitative health data to facilitate timely, accurate diagnosis across various medical domains [2–4].

Oral cancer remains a major global health concern, with approximately 377,000 new cases diagnosed annually and an estimated 177,000 deaths [5,6]. The high mortality rate is largely attributable to late-stage diagnosis, highlighting the urgent need for early detection and timely intervention. Despite this, limited awareness and access to healthcare resources often prevent individuals from recognising early symptoms and seeking appropriate care [7]. In this context, chatbots represent a promising tool for early diagnosis due to their user-friendly interfaces and widespread availability on mobile devices.

The study investigates the application of artificial intelligence for oral cancer detection through an interactive chatbot that performs real-time symptom assessment via automated image classification. The system combines convolutional neural networks (CNNs) with class activation mapping (CAM) to provide visual interpretability, and enhance user trust in model outputs. Unlike conventional tools, the platform offers web and mobile accessibility, an intuitive interface, and personalised guidance. While chatbots have been applied in mental health, chronic disease management, and patient education [8,9], their use in oncology remains limited due to the absence of real-time visual analysis and interactivity [10]. The novelty of this work lies in adapting established CNN architectures for oral cancer image classification within a patient-facing chatbot environment, facilitating more informed user engagement. This study clarifies its methodological orientation and contributions rather than offering an extended comparison of existing chatbot applications. It presents an exploratory implementation of deep learning-based image classification within a conversational interface, aimed at assessing feasibility rather than clinical-grade performance. The system demonstrates a lightweight, patient-facing design that integrates convolutional neural networks, natural language processing, and class activation mapping. Emphasis is placed on deployment in resource-constrained settings, effective human–computer dialogue for image submission and health education, and the broader contribution of linking algorithmic performance with patient-centred interaction.

Despite advances in healthcare technology, digital tools for early symptom recognition and public education in oral cancer remain scarce. Existing applications often provide static content with minimal interactivity, lacking features such as image-based analysis or adaptive user feedback [11,12]. These limitations reduce their effectiveness, particularly for underserved populations facing barriers to care [13]. This underscores the need for a specialised, interactive solution that delivers tailored information, enhances symptom recognition, and provides meaningful support to individuals at risk of oral cancer. Chatbots offer a promising means of supporting public health, particularly in settings with limited access to healthcare professionals. Accessible via mobile devices, they can provide timely, low-cost guidance to underserved populations. Evidence shows that conversational agents enhance health literacy, promote early symptom recognition, and encourage engagement in preventive behaviours [14]. By delivering interactive and personalised feedback, chatbots advance beyond static health applications, aligning with digital health priorities of scalability, inclusivity, and user-centred design.

Recent advances in artificial intelligence and deep learning have significantly improved diagnostic accuracy and patient outcomes in various medical domains. The authors of [15] demonstrated the effectiveness of convolutional neural networks in supporting dermatological diagnoses, particularly for skin cancer detection. Similarly, Ref. [16] high-

lighted the role of natural language processing in healthcare chatbots to facilitate relevant information delivery and enhance patient engagement. However, these innovations remain underutilised in oral cancer diagnosis, where existing solutions continue to rely on conventional methods or generic mobile applications that often lack interactivity and real-time image analysis capabilities [17].

We employ two publicly available datasets [18,19] to train and evaluate CNN models, including DenseNet, and custom designs, with emphasis on balancing sensitivity and specificity. This work contributes to narrowing the gap between technological advancement and clinical practice by supporting early detection and encouraging timely medical consultation. It is important to note that the tool is not diagnostic; rather, it serves as a screening aid that prompts users to seek professional care when potential risks are identified.

This study is guided by the following research questions:

- RQ1: Which artificial intelligence model demonstrates the highest performance in detecting oral cancer based on accuracy, sensitivity, and specificity?
- RQ2: To what extent does the best-performing model exhibit potential for clinical integration and real-world applicability in the early detection of oral cancer?

These research questions establish a clear framework for evaluating model performance and assessing clinical applicability. By addressing them, the study advances the integration of AI-based image classification into healthcare practice, ensuring both methodological rigour and clinical relevance. It contributes to closing the gap between technical performance and practical deployment in the early detection of oral cancer.

The remainder of this paper is structured as follows. Section 2 reviews the existing literature, identifying recent advances in chatbot technology and current limitations in tools for oral cancer detection. Section 3 outlines the materials and methods, including the chatbot's architecture, datasets, and the deep learning models employed for image classification. Section 4 presents the performance outcomes, focusing on the system's ability to differentiate between cancerous and non-cancerous lesions and to educate users on relevant symptoms and risk factors. Section 5 examines the implications of the findings, emphasising the role of chatbots in enhancing diagnostic pathways, patient engagement, and public health strategies. Finally, Section 6 summarises the key contributions and suggests directions for future research and development of AI-based healthcare solutions.

2. Related Work

This section reviews existing chatbot technologies in the healthcare domain, focusing on their capabilities, limitations, and integration with artificial intelligence (AI). While AI-powered chatbots have been deployed in various medical contexts, their application in oral cancer detection remains scarce. Most existing systems lack interactive features, personalised guidance, and image-based diagnostic capabilities—elements that are essential for early detection and user engagement. These limitations highlight the need for specialised solutions that combine conversational AI with visual analysis to improve patient support and early symptom triage.

Healthcare professionals increasingly adopt chatbots to educate patients, monitor symptoms, and support the management of chronic diseases [20,21]. However, most solutions related to oral cancer remain limited to static information delivery, lacking interactive features such as real-time image analysis or personalised guidance [17,22].

Advancements in artificial intelligence have significantly enhanced chatbot development. Natural language processing (NLP) enables chatbots to understand and generate human language, allowing them to engage in purposeful and coherent conversations with their patients [23]. Furthermore, machine learning algorithms strengthen these capabilities by learning from user interactions, which enables chatbots to adapt and improve over time.

This continuous learning process increases the effectiveness of addressing individual user needs. Frameworks such as spaCy, NLTK, and the GPT-3 API have been widely adopted in general-purpose conversational agents [16,24]. However, the adoption of these technologies in specialized domains such as oral cancer remains limited. This underutilization presents a clear opportunity to leverage AI advancements in underserved communities, fostering more inclusive and targeted healthcare solutions.

Chatbots have also shown promise in diverse healthcare domains, including chronic disease management and cancer care, where they have been used to improve patient outcomes, provide coping strategies, and reduce anxiety [21,25,26]. Despite these benefits, their application to oral cancer remains rare, with few systems offering real-time diagnostic support or interactive patient engagement.

Recent studies have reported promising results for oral cancer detection using alternative architectures, including ResNet, EfficientNet, and more recently, transformer-based models that leverage attention mechanisms for improved feature extraction [27]. These approaches often achieve higher accuracy and robustness compared to conventional CNN variants, particularly when trained on larger and clinically validated datasets. In this study, we focused on DenseNet and Inception as representative and widely used architectures that balance performance with computational efficiency, making them suitable for an initial proof-of-concept prototype. Future work will extend this comparison to incorporate more advanced models, thereby enhancing both the comprehensiveness and competitiveness of the system.

Advancements in NLP and convolutional neural networks (CNNs) have significantly expanded chatbot functionality. However, their application to oral cancer detection remains limited [16]. Current systems fail to integrate medical imaging, which is a critical component for diagnosing visual conditions, such as oral cancer. Although several tools support general cancer care [26], they lack real-time analytical capabilities, and do not address lesion detection. Class Activation Mapping (CAM) offers a promising solution by providing visual interpretability, which strengthens user trust and enhances the system's clinical utility for both patients and healthcare professionals.

Despite these advances, several challenges hinder the effective deployment of AI-driven chatbots in clinical settings. Key issues include diagnostic accuracy, data privacy, integration with electronic health records (EHRs), and user trust [28–30]. These challenges are particularly acute in underserved populations, where access to digital tools is limited. Most existing applications offer static content without personalised support or visual analysis, limiting their effectiveness for oral cancer triage [31,32]. To address these shortcomings, this study introduces a functional, interactive system that integrates CNN-based image classification with real-time conversational support.

3. Materials and Methods

This section describes the methodology used to design and implement the oral cancer chatbot, including system architecture, data preparation, model development, and evaluation procedures. The platform integrates convolutional neural network (CNN) variants to classify oral lesion images as cancerous or non-cancerous. Key preprocessing steps—such as image resizing, noise reduction, and dataset partitioning—were applied to ensure consistency and robustness. Hyperparameter tuning procedures are also presented. All methodological decisions were aligned with the study's aim of enhancing diagnostic accuracy and accessibility.

3.1. Chatbot Design Considerations

The oral cancer chatbot is designed to maximise usability, accessibility, and reliability for individuals at risk. Its architecture comprises three core components: a user-facing frontend, a backend for secure data management, and Python-based processing engines. The system adheres to ethical guidelines by safeguarding user data and explicitly stating its non-diagnostic function, directing users to seek professional medical advice when appropriate. The frontend supports intuitive interaction, while the backend manages communication and storage. The NLP module processes user input, and the image analysis engine performs real-time symptom evaluation. Figure 1 depicts the system workflow and component integration. The chatbot follows an interactive flow that combines image submission with symptom inquiry. Users receive brief instructions for capturing and uploading an oral image via smartphone, followed by structured questions on lesion characteristics and risk factors. The neural network then processes the image in real time and returns a probability-based output with tailored educational messages, fostering informed engagement and encouraging timely consultation with healthcare professionals.

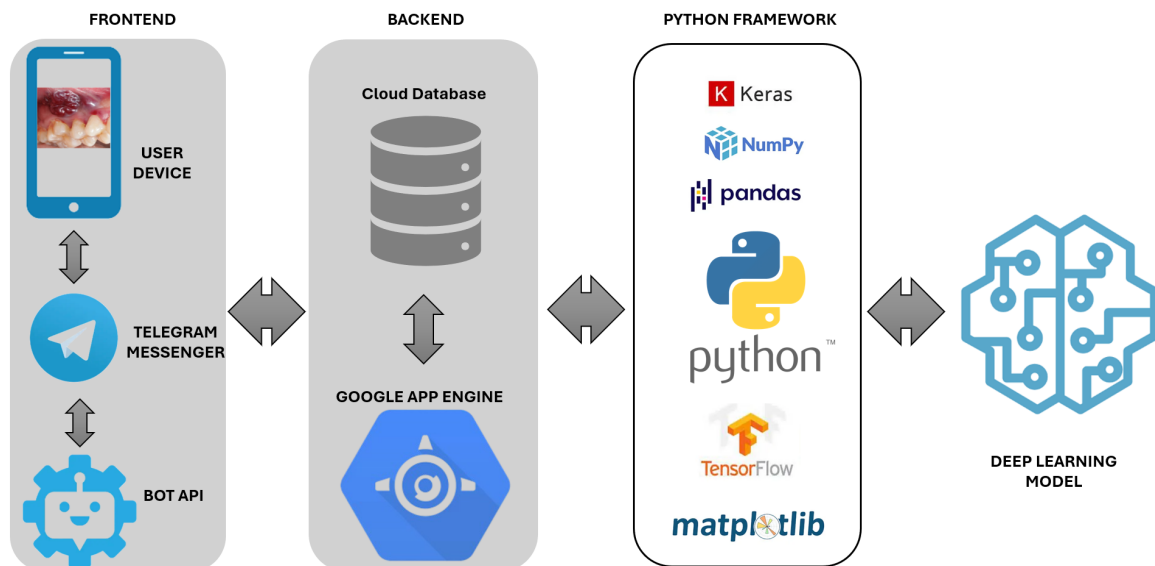


Figure 1. System architecture of the oral cancer chatbot, illustrating the interactive workflow between the Telegram interface, user inputs (image uploads and symptom questions), and the backend modules. The backend integrates AI-driven image analysis using convolutional neural networks (CNNs) with Class Activation Mapping (CAM) for visual interpretability, and combines these outputs with conversational logic to provide users with probability scores, highlighted lesion regions, and tailored educational guidance.

The chatbot was designed to optimise usability, accessibility, and reliability for individuals at risk of oral cancer. Its architecture includes a user-facing frontend, a backend for secure data handling, and Python-based inference and processing modules. The system adheres to ethical standards by safeguarding user data and clearly stating its non-diagnostic function, directing users to seek professional medical advice when necessary. The frontend, implemented on the Telegram platform, supports intuitive interactions for image-based symptom assessment and health education. The probability values presented to users represent model-derived confidence scores rather than prevalence-adjusted predictive values. These outputs indicate the relative certainty of the neural network based on training data and should not be interpreted as reflecting the true likelihood of disease in a screening population. It is developed in Spanish and aligns with Technology Readiness Level 2 (TRL2) standards [33]; it employs straightforward commands and interactive buttons to guide

users. Accessibility features, including screen reader compatibility, simplified navigation for users with cognitive or motor impairments, and visual aids, enhance usability across diverse populations and promote wider adoption. The chatbot also guides users on image capture by advising on lighting, focus, and camera positioning to approximate the quality of training data. Despite such instructions, variability in patient-generated images remains inevitable. Future deployment must include structured guidelines and training resources to reduce inconsistency and improve compatibility with clinically relevant imaging standards.

The backend infrastructure ensures secure, scalable, and efficient data handling, fully compliant with the General Data Protection Regulation (GDPR) [34]. Built on Google App Engine and Cloud Storage [35], it supports real-time image analysis and responsive user communication. The architecture integrates AI-driven image processing with Telegram's API, enabling dynamic request management, reliable data storage, and secure interaction, all of which are essential for effective deployment in healthcare settings.

Python 3.13.7 underpins the system's backend, enabling seamless machine learning model implementation, data processing, and cloud service integration. Its simplicity and extensive library support facilitated rapid prototyping and model iteration. TensorFlow and Keras supported neural network development, while NumPy and Pandas ensured efficient data management. Matplotlib 1.5.x provided visualisation capabilities [36], together forming a robust and scalable framework suitable for clinical-grade applications.

3.2. Deep Learning Models

The system employs a range of convolutional neural network (CNN) architectures, including DenseNet121, DenseNet169, DenseNet201, Inception-v3, and a custom Generic CNN, to support automated classification of oral lesions. DenseNet models were selected for their densely connected architecture, which promotes efficient gradient propagation and feature reuse—key characteristics for enhancing diagnostic performance in medical imaging. Inception-v3 introduces multi-scale convolutional modules that capture spatial features at various resolutions, improving lesion detection in cases of morphological variability. The Generic CNN, serving as a baseline model, provides a reference for assessing performance improvements achieved by deeper architectures. To enhance model transparency, the system integrates Class Activation Mapping (CAM) [37], which not only identifies the image regions contributing most to each prediction but also transmits these highlighted regions back to the chatbot interface. This enables the chatbot to provide users with a visual explanation of the model's decision, thereby improving interpretability and trust.

- **DenseNet121, DenseNet169, and DenseNet201 [38]:** These variants differ in depth (121, 169, and 201 layers, respectively) and feature dense connectivity, where each layer is linked to all subsequent layers. This architecture promotes feature reuse and stable gradient flow, allowing the network to learn compact and discriminative representations, which is well suited for high-precision classification tasks in healthcare.
- **Generic CNN [39]:** A standard convolutional neural network is included as a baseline to evaluate the added value of more complex models. It learns spatial features directly from input data and offers a comparative reference for assessing model performance.
- **Inception-v3 [40]:** It utilises multi-scale convolutional filters within its modules to capture spatial features at varying resolutions, making it particularly effective in handling the morphological variability of oral lesions.

Deep learning models underpin the system by enabling rapid, image-based assessments and supporting early detection of oral cancer. The architectural progression from standard convolutional neural networks (CNNs) to advanced designs such as DenseNet and InceptionV3 demonstrates key innovations in network efficiency and diagnostic accuracy. Traditional CNNs, which rely on sequential layer stacking, often face challenges

such as vanishing gradients and limited feature reuse. DenseNet addresses these limitations by establishing direct connections between all layers, enhancing feature propagation and improving learning efficiency. InceptionV3 further contributes by employing multi-scale filter operations and parallel convolutional paths, enabling more effective extraction of discriminative features from medical images. These enhancements make DenseNet and InceptionV3 well suited for robust, interpretable image-based diagnosis in clinical decision-support systems. Each model underwent systematic hyperparameter optimisation, including tuning of learning rates, batch sizes, and architectural configurations, to enhance performance on the selected datasets. These architectures were chosen for their proven accuracy in medical image analysis and their suitability for developing a robust and interpretable tool for oral cancer diagnosis.

3.3. Oral Cancer Data Preprocessing

This section shows the complete pipeline of the DL classification. Figure 2 shows the process.

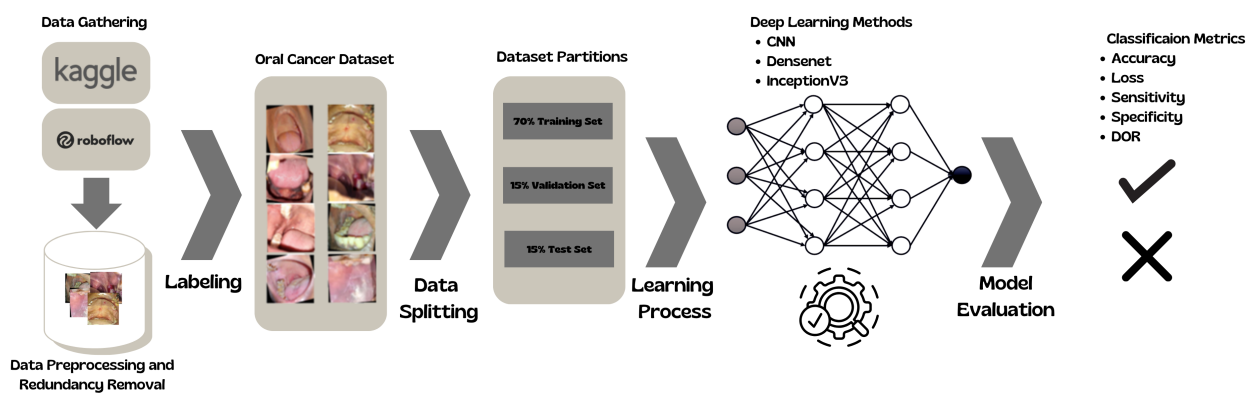


Figure 2. Deep learning classification data pipeline.

Figure 3 illustrates the overall workflow of the proposed chatbot, highlighting the interaction between users, the neural network pipeline, and the outputs presented to patients.

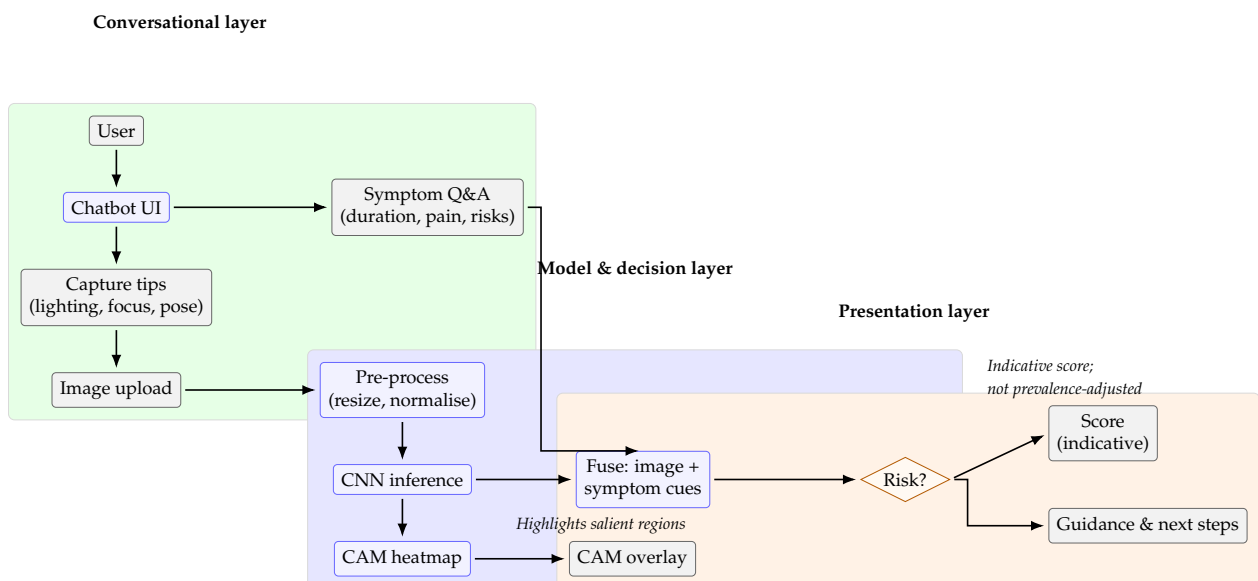


Figure 3. Workflow of the proposed chatbot. The conversational layer collects oral images and symptom information; the model and decision layer processes inputs through pre-processing, CNN inference, CAM generation, and risk assessment; and the presentation layer delivers an indicative probability score, a CAM-based visual explanation, and tailored guidance.

Datasets from Kaggle [18] and Roboflow [19] were selected to ensure a diverse and representative image set for robust model training and evaluation. The Kaggle dataset includes 500 cancerous and 450 non-cancerous oral lesion images, while the Roboflow dataset contributes an additional 1700 images, enhancing variability and generalisability. Ethical standards were upheld by excluding personally identifiable information (PII) [41], and all data usage complied with the respective licensing terms. A summary of the datasets is presented in Table 1. Both datasets were selected primarily for their accessibility, balanced composition, and suitability for prototyping deep learning models. They provided sufficient annotated images to support initial training and evaluation of convolutional architectures, enabling the proof-of-concept development of the chatbot. However, these datasets do not reflect the epidemiological profile of real-world screening, where prevalence is low and early-stage cases are common. Moreover, they lack clinical context, such as patient history and oncologist-confirmed diagnoses, which limits their representativeness. Future work will therefore focus on validation with clinician-curated datasets containing expert-verified ground truth to ensure stronger alignment with clinical scenarios and translational relevance.

Table 1. Composition of training datasets used for oral cancer image classification.

Dataset	Cancer Images	Non-Cancer Images
Kaggle Dataset	500	450
Roboflow Dataset	850	850
Total	1350	1300

The image preprocessing pipeline was designed to standardise and optimise the datasets for deep learning applications. Images from Kaggle and Roboflow were resized to 256×256 pixels to ensure uniform input dimensions and facilitate efficient training. A 2-pixel black border was added to mitigate edge effects during convolutional operations. A mild blur using a 2×2 kernel was applied to reduce noise and enhance generalisability by suppressing high-frequency artefacts.

To eliminate redundancy, a duplicate detection process was implemented using root mean square (RMS) error calculations. Images were converted to NumPy arrays, and pairs with an RMS error below a threshold of 3 were flagged as duplicates and removed. The resulting dataset, comprising uniquely preprocessed images, was stored in organised directories for training and evaluation. This systematic preprocessing enhanced data quality, reduced bias from repeated samples, and improved model reliability.

The dataset was randomly shuffled and partitioned into training (70%), validation (15%), and testing (15%) subsets using NumPy's `np.split` function to ensure unbiased model evaluation. Each subset was stored in dedicated directories for structured access during training and performance assessment. This systematic division ensured a balanced representation across all phases of model development.

Figures 4 and 5 illustrate representative chatbot interaction scenarios that support user engagement and the dissemination of reliable oral cancer information. The first example demonstrates how the system delivers probability-based feedback and guidance following an image submission. The second highlights the use of Class Activation Mapping to visualise regions of clinical interest, enhancing interpretability and promoting informed consultation. Together, these interactions reflect a user-centred design that combines clarity, accessibility, and contextual support to address diverse oral health queries.

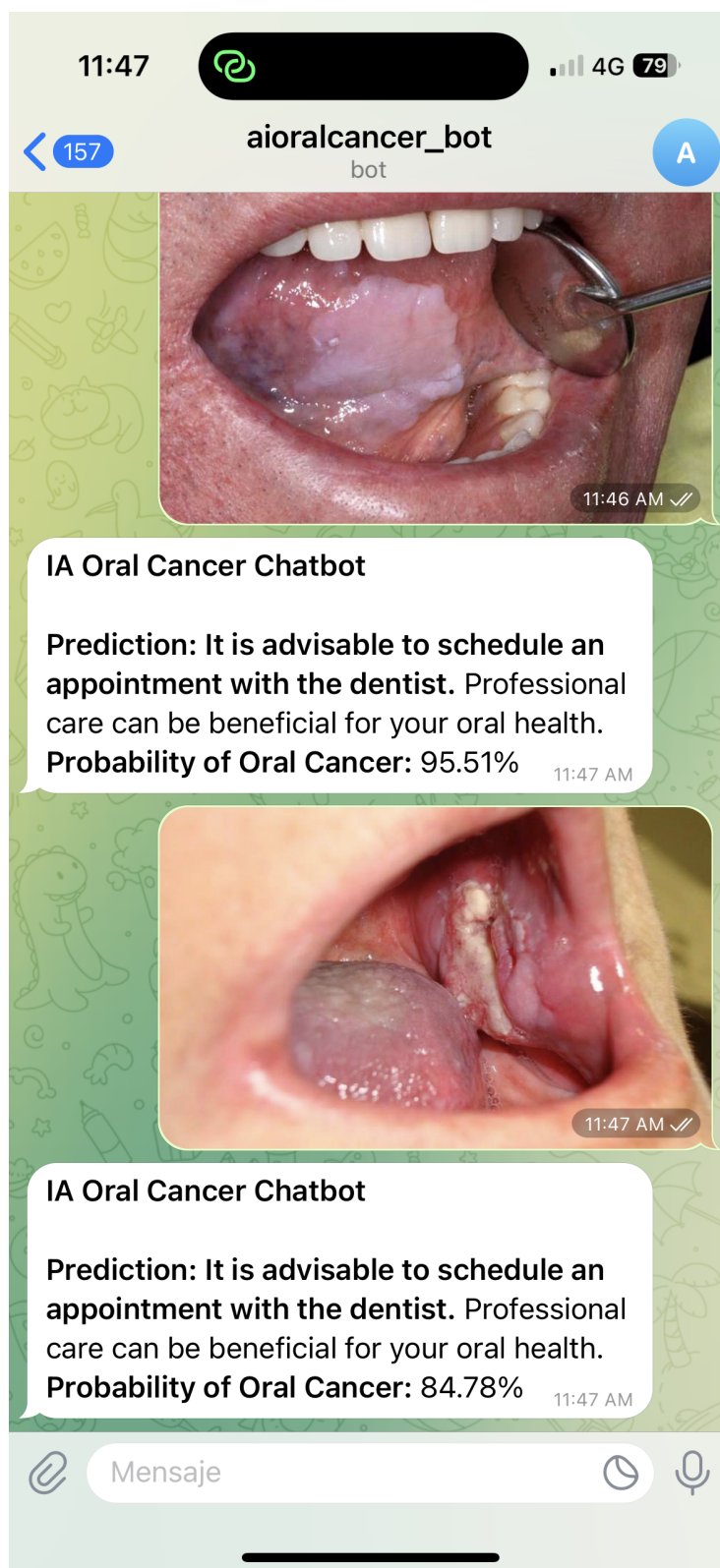


Figure 4. Chatbot interface displaying the predicted probability of oral cancer alongside tailored guidance. The output includes both numerical feedback and educational recommendations, helping users interpret the system's results.

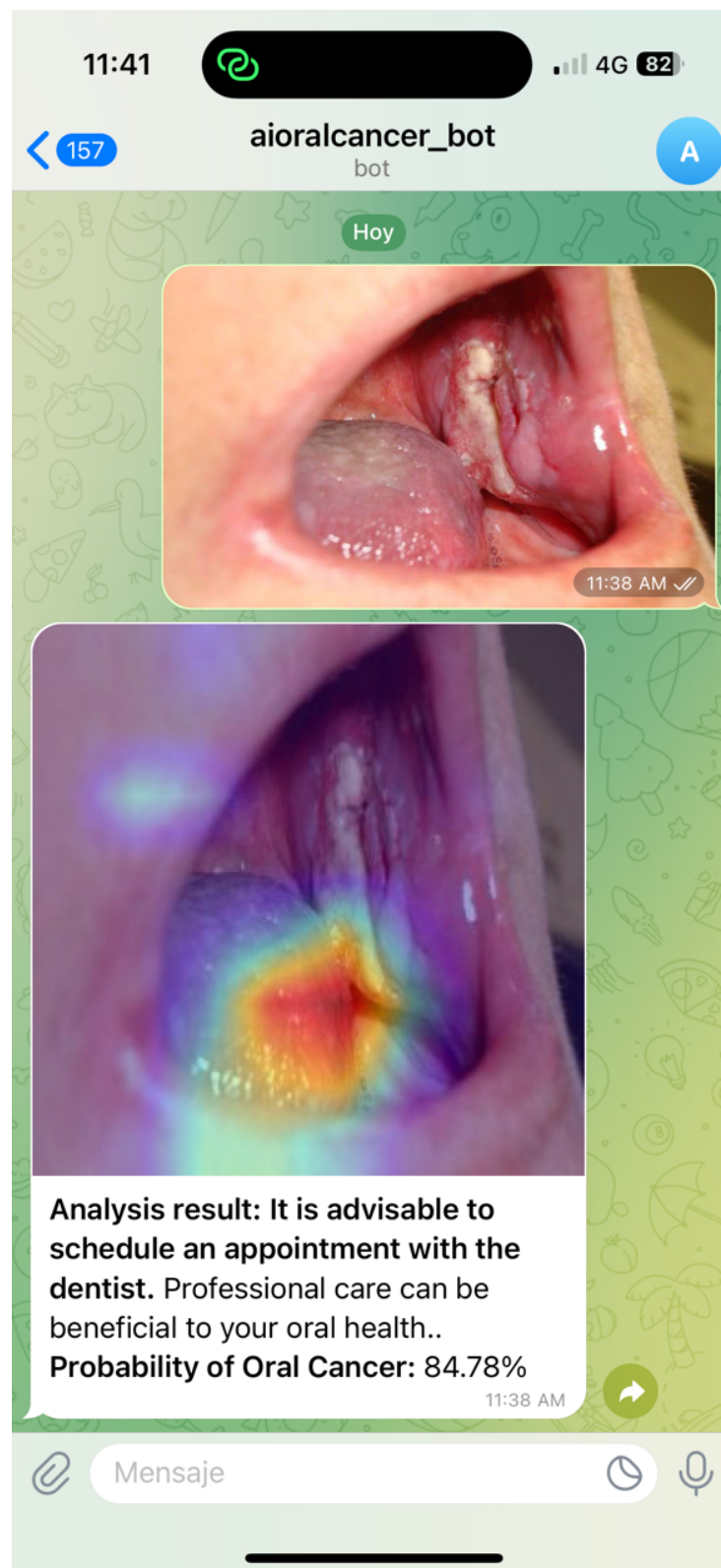


Figure 5. Class Activation Mapping (CAM) highlighting the relevant region of the uploaded image. This CAM output is integrated into the chatbot interface to provide patients with a visual explanation of the model's prediction, supporting interpretability and informed decision-making.

3.4. Development Process and Implementation

The chatbot was developed using an agile methodology, enabling iterative improvement and seamless integration of user feedback. The development process began with a requirement analysis informed by interviews with patients, healthcare professionals, and

caregivers, identifying critical system features. A benchmarking review of existing digital health tools guided best practices and revealed functional gaps, informing the scope and design of the platform. Core capabilities, including symptom assessment, health education, and support services, were prioritised to meet user needs.

The prototyping phase involved designing wireframes and mockups to define the user interface, followed by a functional prototype incorporating basic conversational abilities. Usability testing with a targeted user group refined the interface and confirmed alignment with user expectations.

Subsequent development followed sprint cycles, allowing feature implementation guided by prioritised feedback. The final phase involved comprehensive testing, including alpha testing for system stability, beta testing with a wider user base, and adjustments based on real-world interactions. Functional, usability, and security testing ensured system reliability and compliance with data protection standards.

This structured methodology produced a secure and user-centred chatbot to support early oral cancer detection and patient education. Future enhancements will target expanded functionality, improved usability, and responsiveness to evolving healthcare needs.

Methodological decisions, from dataset selection to model architecture, aligned with the study's objective of improving diagnostic accuracy and accessibility. The framework ensured transparency and reproducibility while addressing the research questions with clinical relevance.

To address RQ1, a comparative analysis of deep learning models will be conducted using two heterogeneous datasets. Each model sustained systematic training, hyperparameter optimisation, and evaluation using accuracy, sensitivity, and specificity. For RQ2, the best-performing model will be benchmarked against established diagnostic tools to assess its practical applicability. This methodology ensured the study objectives will be met through a rigorous and clinically meaningful approach.

3.5. Hardware and Computational Performance

To evaluate the practical feasibility of the proposed system, we report details of the hardware used during training and inference. All experiments were conducted on a workstation equipped with an AMD Ryzen 7 5800U processor, Advanced Micro Devices, Inc., Santa Clara, CA, USA (1.90 GHz, 16 GB RAM). Training was performed using TensorFlow 2.9 and Keras 2.9 frameworks, with batch sizes of 32 and image resolution of 256×256 pixels.

As no dedicated GPU was available, model training was computationally demanding, particularly for deeper architectures such as DenseNet201, which required extended runtimes to converge. While inference was executable on CPU hardware, detailed measurements of latency, memory consumption, and power usage were not performed. These factors remain essential for practical deployment and will be addressed in future evaluations using both mobile and embedded platforms.

3.6. Evaluation Metrics

To evaluate model performance comprehensively, we applied multiple classification metrics: loss, accuracy, specificity, sensitivity, diagnostic odds ratio (DOR), and Cohen's Kappa. These metrics collectively assess prediction accuracy, classification error, discriminative ability, and agreement beyond chance. Table 2 defines each metric and outlines its formula. All metrics derive from standard classification outcomes: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). TP denotes correctly identified positive cases, FP refers to incorrect positive predictions, TN corresponds to correctly identified negative cases, and FN indicates missed positive cases. These values serve as the foundation for calculating each performance metric.

Table 2. Performance metrics used for model evaluation.

Metric	Description and Formula
Loss (%)	Measures the prediction error between model outputs and true labels. It guides training by penalising incorrect predictions.
Accuracy (%)	Indicates the overall proportion of correct predictions: $\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100$
Specificity (%)	Proportion of correctly identified negatives: $\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100$
Sensitivity (%)	Also called recall; measures the proportion of correctly identified positives: $\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100$
DOR	Diagnostic odds ratio is a single indicator of test performance: $\text{DOR} = \frac{\text{TP} \times \text{TN}}{\text{FP} \times \text{FN}}$
Cohen's Kappa	Higher values indicate better discriminative performance. Measures agreement between predicted and actual labels, adjusted for chance: $\kappa = \frac{p_o - p_e}{1 - p_e}$ where p_o is observed agreement and p_e is expected agreement by chance.

3.7. Hyperparameter Tuning for Model Optimisation

Effective hyperparameter tuning is critical for optimising the performance of machine learning models. By systematically adjusting key parameters, such as learning rates, batch sizes, and the number of layers, the models can achieve improved accuracy, sensitivity, and specificity. This process ensures that the models are not only well fitted to the training data but also generalise effectively to unseen datasets. Table 3 summarises the key hyperparameters tested during the tuning process and the configurations selected for each model.

Table 3. Summary of model architectures. All models were trained with a learning rate of 0.001, batch size of 32, and 50 epochs.

Model	Architectural Details
DenseNet-121	Growth rate: 32; Layers: 121
DenseNet-169	Growth rate: 32; Layers: 169
DenseNet-201	Growth rate: 32; Layers: 201
CNN	Filter size: 3×3 ; Number of filters: 64
Inception-v3	Filter sizes: 1×1 , 3×3 , 5×5 ; Pooling: Max/Average

The hyperparameters were selected based on established practices in medical imaging and validated through preliminary experiments. This configuration consistently achieved stable convergence and generalisable performance, offering a balanced trade-off between training efficiency and classification accuracy.

3.8. Code and Data Availability

To support reproducibility and future research, all code and preprocessing scripts developed for this study will be made publicly available in an open access GitHub repository. The repository contains the full implementation of the chatbot framework, including data preprocessing pipelines, model training scripts, and inference routines. Sensitive clinical data are not shared; however, anonymised code and workflows are provided to enable replication and extension of the proposed approach. The repository is accessible at the following link: <https://github.com/eduardonavarrol/Chatbot-Detection-Oral-Cancer> (accessed on 16 May 2025).

4. Results

This section presents the comparative performance of the evaluated models using six key metrics: accuracy, loss, sensitivity, specificity, diagnostic odds ratio (DOR), and Cohen's Kappa. Sensitivity is emphasized due to its critical role in early oral cancer detection, while specificity captures the ability to avoid false positives. The inclusion of Cohen's Kappa offers insight into agreement beyond chance.

4.1. System Performance

Evaluating the system performance is essential for determining its effectiveness in accurately detecting oral cancer and delivering reliable user support. Metrics such as accuracy, loss, sensitivity, and specificity provide a comprehensive assessment of the models' predictive capabilities. These indicators reveal both strengths and limitations, informing model refinement and supporting validation in real-world applications. Table 4 presents a detailed comparison of the performance metrics across the evaluated models.

Table 4. Comparative performance of five deep learning models across six evaluation metrics for oral cancer classification.

Method	Accuracy (%)	Loss	Specificity (%)	Sensitivity (%)	DOR	Kappa
CNN	65.06 (7.13)	0.62 (0.04)	57.05 (20.31)	65.90 (7.73)	7.62 (11.10)	0.230
DNet201	71.45 (11.51)	0.59 (0.23)	48.44 (27.17)	66.21 (12.24)	4.58 (4.89)	0.147
DNet169	73.36 (8.69)	0.70 (0.21)	50.81 (12.08)	67.05 (9.20)	4.14 (3.61)	0.179
DNet121	68.94 (8.07)	0.75 (0.24)	42.92 (14.48)	63.01 (7.43)	2.58 (3.15)	0.059
INCV3	77.65 (2.98)	0.47 (0.06)	77.68 (8.24)	78.02 (11.29)	20.67 (12.44)	0.557

The comparative analysis of model performance reveals InceptionV3 as the most effective architecture for oral cancer detection within this study. It achieved the highest mean accuracy, along with the top sensitivity and specificity. Furthermore, it recorded the highest DOR and a substantial Cohen's Kappa score of 0.557, indicating moderate agreement beyond chance and suggesting a balanced capability in detecting both positive and negative cases. These results highlight the robustness of InceptionV3 in managing both class types effectively. Given its superior performance, InceptionV3 was integrated into the chatbot's diagnostic module to support accurate and responsive image-based assessments. This integration enhances the chatbot's ability to provide personalised and clinically meaningful feedback, improving user trust and reinforcing the system's potential as a reliable digital health assistant.

DenseNet201 and DenseNet169 demonstrated moderate performance overall. Although DenseNet201 achieved the second-highest mean accuracy, it showed limited specificity and a low Cohen’s Kappa score, indicating only slight agreement beyond chance. DenseNet169 ranked similarly in accuracy but also struggled to differentiate between classes, as reflected by its modest sensitivity and specificity and weak agreement. These findings suggest that while DenseNet architectures exhibit strong feature extraction, their practical deployment may require further tuning. DenseNet121 and the custom CNN model delivered the weakest predictive agreement, with Kappa scores of 0.059 and 0.230, respectively. DenseNet121 recorded the lowest specificity and DOR, indicating a high rate of false positives. Although the CNN model achieved more consistent sensitivity and accuracy, its low specificity also limits its clinical utility. Based on these performance values, DenseNet201 and DenseNet169 could be integrated into the chatbot as intermediate-tier diagnostic aids, offering reasonable accuracy while flagging potential risk cases. However, due to their lower specificity and weak agreement scores, their outputs should be supplemented with additional screening questions or escalated for clinical review to avoid misclassification. Figure 6 presents a grouped bar chart comparing the performance of five deep learning models—CNN, DenseNet121, DenseNet169, DenseNet201, and InceptionV3—across three key evaluation metrics.

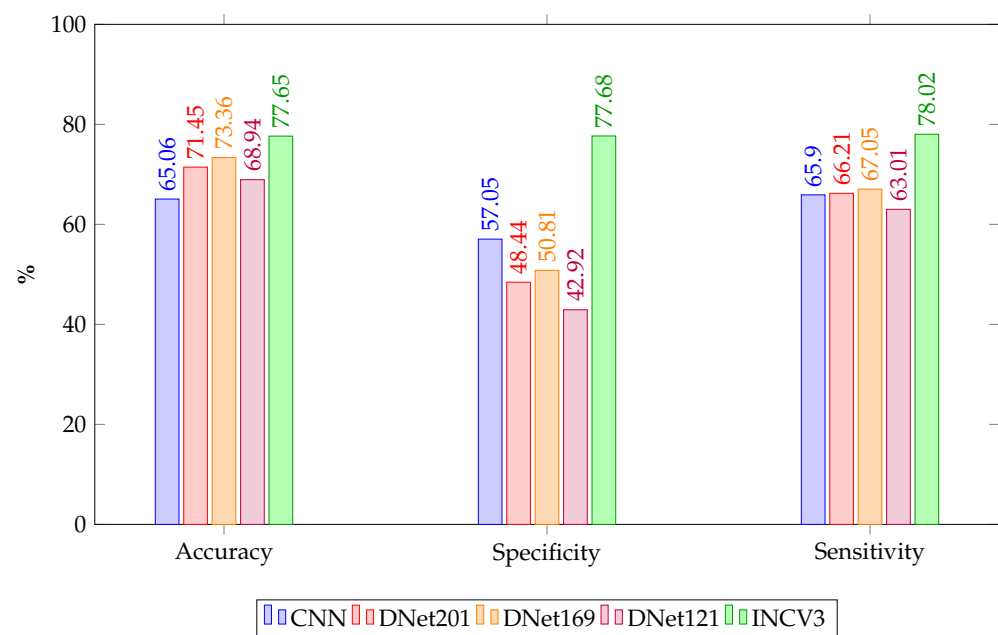


Figure 6. Performance comparison of CNN, DenseNet variants, and InceptionV3 across accuracy, sensitivity, and specificity.

Figure 7 presents a comparison of the diagnostic odds ratio (DOR) across the five evaluated models. InceptionV3 clearly stands out with the strongest overall discriminative ability, making it the most reliable model for distinguishing between positive and negative cases. The CNN model also shows solid performance despite its simpler architecture. In contrast, the DenseNet variants perform moderately, with DenseNet121 displaying the lowest discriminative strength. These results support the suitability of InceptionV3 as the core diagnostic engine within the chatbot system.

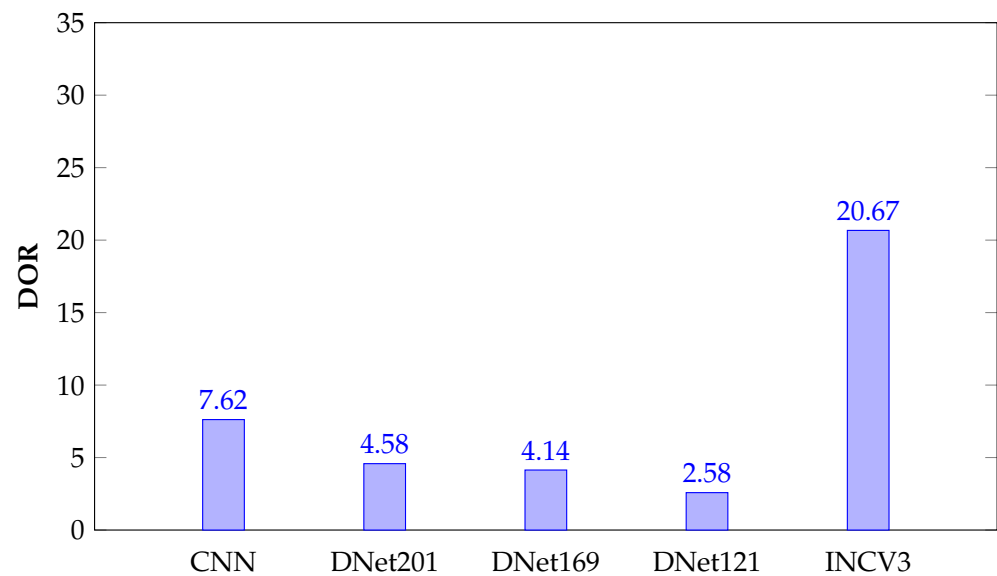


Figure 7. Diagnostic odds ratio (DOR) for each model. InceptionV3 shows the highest discriminative capacity.

4.2. Statistical Analysis

To comprehensively assess the effectiveness of the AI models in oral cancer detection, Table 5 presents key performance metrics, including accuracy, loss, specificity, sensitivity, diagnostic odds ratio (DOR), and Cohen's Kappa, each reported with their respective 95% confidence intervals. These metrics provide a statistically rigorous evaluation of each model's diagnostic reliability and highlight their potential for clinical deployment.

Table 5. Performance metrics with 95% confidence intervals for all evaluated models.

Model	Accuracy	Loss	Specificity	Sensitivity	DOR	Kappa
CNN	(61.15, 68.96)	(0.58, 0.66)	(53.21, 60.89)	(62.17, 69.63)	(3.99, 11.25)	(0.204, 0.256)
DenseNet201	(67.31, 75.59)	(0.36, 0.82)	(40.19, 56.69)	(60.46, 71.96)	(3.66, 5.50)	(0.104, 0.190)
DenseNet169	(69.99, 76.73)	(0.49, 0.91)	(46.74, 54.88)	(63.34, 70.76)	(3.82, 4.46)	(0.141, 0.217)
DenseNet121	(64.46, 73.42)	(0.51, 0.99)	(37.63, 48.21)	(59.79, 66.23)	(2.68, 3.08)	(0.036, 0.082)
InceptionV3	(74.67, 80.63)	(0.41, 0.53)	(72.79, 82.57)	(71.86, 84.18)	(18.23, 23.11)	(0.537, 0.577)

Confidence interval analysis provided insight into model stability and generalisation. InceptionV3 exhibited the narrowest confidence intervals and highest scores across all core metrics, reinforcing its robustness and discriminative capacity. DenseNet169 achieved competitive accuracy, but with slightly broader intervals and lower specificity. DenseNet201 performed moderately well but showed more variable results. DenseNet121 and the custom CNN model demonstrated weaker performance, with broader confidence intervals and lower agreement beyond chance, indicating reduced reliability in clinical settings.

Overall, InceptionV3 and DenseNet169 demonstrated the most balanced performance, supported by narrow confidence intervals and robust diagnostic scores. InceptionV3's consistent accuracy and interpretability make it especially well suited for integration into AI-driven oral cancer screening workflows. In contrast, CNN and DenseNet121 presented significant limitations in sensitivity, specificity, or agreement metrics, highlighting their limited suitability for high-stakes medical applications without further optimisation.

5. Discussion

The results reveal a performance trade-off between sensitivity and specificity, emphasising the importance of selecting models based on specific clinical use cases. DenseNet201

provides a balanced option for general screening scenarios, whereas InceptionV3 demonstrates superior discriminative power, minimising false positives. These findings support a hybrid integration strategy, where multiple CNN architectures are employed within chatbot systems to balance early risk detection with diagnostic precision. Such tools, when embedded into user-facing platforms, have the potential to empower individuals through personalised, image-based assessments and timely access to educational guidance, thereby promoting early intervention and improved health outcomes [42–44]. From a computational standpoint, InceptionV3 outperforms shallower models like CNNs and DenseNet121 in both accuracy and robustness, exhibiting narrower confidence intervals, higher DOR, and moderate Cohen's Kappa. Its parallel convolutions capture multi-scale features, improving lesion detection. Clinically, its high sensitivity and specificity help minimise false negatives and overdiagnosis, enhancing early-stage oral cancer detection while reducing patient anxiety and unnecessary interventions. Although the chatbot applies deep learning for oral cancer detection, it is not intended as a population-level screening tool, which requires very high sensitivity and negative predictive value. The performance reported here falls short of clinical thresholds and should therefore be viewed as exploratory. The system is better positioned as a pre-consultation aid that raises awareness, encourages timely medical visits, and guides individuals toward professional evaluation, serving as an assistive mechanism rather than a diagnostic substitute.

Integrating AI-based tools into clinical workflows can reduce diagnostic delays and improve outcomes, especially when models are validated with confidence intervals and supported by interpretable methods like Class Activation Mapping. Chatbots enhance access to vital health information, promoting equity and informed decisions in oral cancer care [45]. Ensuring ethical deployment requires strict data privacy measures, including GDPR-compliant encryption, to foster user trust. A human-in-the-loop framework maintains clinical oversight while leveraging AI efficiency within clearly defined boundaries.

Another important limitation relates to the representativeness of the datasets used. While the Kaggle and Roboflow collections provide sufficient data for prototyping, their composition may not fully capture clinically relevant pathological features, such as the subtle manifestations of early-stage oral cancer. As a result, there remains uncertainty regarding the extent to which the model's classifications correspond to real pathological findings.

While the comparative performance of deep learning models provides valuable insights, the primary contribution of this work lies in demonstrating the integration of image classification and conversational AI within a lightweight chatbot platform. The system is positioned as an auxiliary, non-diagnostic tool that enhances patient engagement, supports early symptom awareness, and offers scalable solutions for resource-constrained environments. This emphasis aligns with the study's stated objective of bridging algorithmic development with practical healthcare applications, rather than solely benchmarking models.

Although InceptionV3 performed best among the tested models, overall accuracy and specificity remained below clinical standards and lower than related studies. The high rate of false positives and limited sensitivity render the models unsuitable for decision support. This work should therefore be viewed as a feasibility prototype integrating deep learning within a chatbot framework. Future research must use larger, clinically validated datasets and advanced architectures to reach performance levels required for reliable healthcare deployment. Beyond algorithmic performance, chatbot deployment depends on usability, acceptability, and user trust. These aspects have not yet been formally evaluated. Future research will include structured user studies with patients and clinicians to assess engagement and reliability in real-world settings.

User engagement is vital for effective AI tool adoption. Adaptive conversational interfaces enhance personalisation, health literacy, and early symptom reporting, especially

in underserved areas. Unlike broader AI healthcare applications, such as CNNs for skin cancer [15] or NLP for emotional support [16], this study uniquely integrates CNN-based image diagnostics with a chatbot for oral cancer detection. This novel combination provides real-time, patient-facing support for early detection and education [17].

In response to RQ1, InceptionV3 demonstrated the strongest diagnostic performance across all evaluated metrics, making it the most suitable candidate for integration into AI-powered oral cancer screening tools. Regarding RQ2, its high sensitivity, specificity, and moderate agreement, combined with modular compatibility for chatbot deployment, indicate strong potential for real-world application. The model's interpretability and stability across datasets support its use in clinical decision-support workflows, particularly in resource-constrained environments.

Although the system design was guided by two research questions, the performance of the tested models was insufficient to fully address the second question regarding applicability in real-world conditions. As a result, RQ2 is only partially answered, and its complete resolution will require further research with larger datasets, improved architectures, and comprehensive evaluation of chatbot usability.

Limitations

Several methodological limitations may affect the internal validity of the findings. Class imbalance and limited heterogeneity in image quality across datasets may have influenced model performance. Although metrics such as DOR and sensitivity support construct validity, the absence of clinical ground-truth labels from expert oncologists limits diagnostic precision. Furthermore, reliance on public datasets without patient history or contextual metadata may have reduced the models' ability to generalise beyond image-based patterns. Image preprocessing steps, such as blurring and resizing, though standardised, may have inadvertently suppressed clinically relevant visual features. An additional limitation lies in the use of publicly available datasets (Kaggle and Roboflow), which, while suitable for initial prototyping, do not capture the prevalence patterns or early-stage presentations characteristic of real-world screening populations, nor do they include expert-verified ground truth annotations. In addition, there is an absence of quantitative profiling for inference time, memory usage, and power consumption; these measurements are critical for assessing feasibility on target hardware platforms and will be included in future work.

The findings are also limited by the absence of clinical validation across diverse healthcare settings. The publicly available datasets used do not reflect the full variability in imaging conditions, demographics, or disease presentation observed in real-world practice. As a result, external validity and generalisability remain constrained. The chatbot was not tested in live clinical environments or with clinician oversight, further limiting conclusions about its operational feasibility in screening programmes. Another limitation is that the chatbot currently reports raw model probabilities, which do not account for pretest probability or population-level prevalence; without such adjustment, these values may be misinterpreted by end users and should be regarded only as indicative scores. Additionally, the overall accuracy and specificity of the tested models fall short of clinical standards, making the current system unsuitable for decision support and highlighting the need for validation with larger, expert-annotated datasets.

From a technical perspective, the model is currently limited to static image classification and does not incorporate multimodal inputs such as patient history or symptom narratives. While the system uses Class Activation Mapping (CAM) to enhance interpretability, further development is needed to ensure clinical explainability and user trust. Privacy measures, although compliant with general standards (GDPR), have not yet im-

plemented advanced techniques such as federated learning or local inference to minimise data exposure. Integration with electronic health records and large-scale telemedicine platforms remains a key challenge for future scalability. In addition, another limitation concerns the variability of patient-generated images, which often lack the controlled quality of training datasets. Without clear guidance and standardisation, these inconsistencies may compromise diagnostic reliability in real-world applications. Addressing this challenge will require the development of structured image-capture protocols and patient education strategies to ensure greater consistency with clinically relevant standards.

The choice of DenseNet and Inception for this study was guided by their balance of established performance and computational feasibility, but we acknowledge that alternative architectures such as ResNet, EfficientNet, and transformer-based models may achieve superior results and will be considered in future iterations of the system.

A critical limitation is that the datasets used did not include expert-verified ground truth annotations, which restricts the diagnostic reliability of the findings and reduces confidence in their clinical applicability. The study also did not include evaluations of usability, user experience, or acceptability, which are essential for establishing the system's effectiveness and adoption; these aspects will be systematically addressed in future clinical and patient-centred studies.

Addressing these limitations through clinical trials, expert collaboration, and iterative development will be essential for translating this prototype into a clinically validated and ethically robust screening tool.

6. Conclusions

This study introduces a patient-centred chatbot that integrates convolutional neural networks (CNNs) and natural language processing (NLP) to support early detection, triage, and education related to oral cancer. Designed for accessibility and scalability, the system delivers image-based risk assessments and interactive guidance through multilingual web and mobile platforms. Model performance was evaluated using publicly available datasets, with InceptionV3 achieving the highest scores across all metrics, including diagnostic accuracy, sensitivity, specificity, and diagnostic odds ratio (DOR). The integration of Class Activation Mapping (CAM) further enhances transparency and clinical interpretability.

In direct response to the research questions, InceptionV3 emerged as the most effective classification model (RQ1), demonstrating strong discriminative power and consistent performance. Its modular design, low latency, and robust accuracy support its feasibility for integration into digital screening pathways and real-world applications (RQ2). These capabilities position the chatbot as a promising tool for facilitating early oral cancer risk assessment, particularly in low-resource or underserved settings. In this regard, the platform provides non-diagnostic, assistive feedback that promotes timely consultation with health-care providers. Its lightweight design and interactive interface support use in public health campaigns, telemedicine, and community awareness. By expanding access to preliminary evaluation and personalised education, the chatbot fosters more equitable early detection support without suggesting replacement of formal screening pathways. It will also involve designing and testing structured image-capture protocols with patients, including practical guidance and validation studies, to ensure that real-world inputs align more closely with the characteristics of training datasets. But beyond model evaluation, this study contributes by illustrating how a chatbot architecture can combine deep learning, interpretability techniques, and conversational interaction to provide accessible, non-diagnostic support for oral cancer awareness and early consultation. This positions the platform within broader digital health pathways, where scalability and inclusivity are as important as algorithmic accuracy.

While RQ1 was addressed by identifying InceptionV3 as the best-performing model in this study, RQ2 remains only partially resolved due to limited performance, and is therefore positioned as a direction for future investigation rather than a definitive outcome.

Future work will address current limitations through clinical validation, user-centred design iterations, and integration with federated learning frameworks to strengthen privacy. Planned enhancements include multimodal data integration, longitudinal tracking, and gamified dialogue to boost engagement and health literacy. These steps will support the evolution of the chatbot from prototype to trusted digital health tool. It will also focus on collaboration with clinicians to curate expert-annotated datasets that better reflect early-stage oral cancer cases and real-world prevalence, thereby strengthening the clinical validity and applicability of the chatbot. In addition, it will also include user studies with patients and clinicians to evaluate usability, acceptability, and trust, ensuring that the chatbot is not only technically robust but also aligned with the practical requirements of real-world healthcare adoption. It will prioritise a verification stage in collaboration with clinicians, including validation against expert-annotated clinical datasets, to ensure that the model's outputs are aligned with pathological features observed in practice. Additionally, it will also involve benchmarking the system on resource-constrained hardware platforms, such as smartphones or edge devices, to quantify inference time, memory requirements, and energy consumption, thereby ensuring suitability for deployment in low-resource healthcare settings. Future studies will involve collaboration with oncologists to curate larger datasets with expert-verified ground truth and improved preprocessing pipelines, thereby enhancing both the robustness and the clinical relevance of the results. Finally, it will prioritise a verification stage in collaboration with clinicians, including validation against expert-annotated clinical datasets, to ensure that the model's outputs are aligned with pathological features observed in practice. And, it will also involve benchmarking the system on resource-constrained hardware platforms, such as smartphones or edge devices, to quantify inference time, memory requirements, and energy consumption, thereby ensuring suitability for deployment in low-resource healthcare settings.

Author Contributions: P.O.-A.: writing—original draft, review and editing, G.M.: writing—original draft, review and editing, C.T.: conceptualisation, writing—review and editing, project administration, funding acquisition, G.G.: writing—review and editing, J.P.V.: writing—review and editing, E.N.: conceptualisation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Agencia Nacional de Investigación y Desarrollo of Chile (ANID), and the Centro de Prevención y Control de Cáncer (CECAN) FONDAP 152220002.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We acknowledge the COVID0739 project and the FONDECYT Regular 1201787 project, titled “Multimodal Machine Learning approach for detecting pathological activity patterns in elderlies”, financed by the Agencia Nacional de Investigación (ANID).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	artificial intelligence
CNN	convolutional neural network
DL	deep learning
DOR	diagnostic odds ratio

CAM class activation mapping
NLP natural language processing

References

1. Zagade, A.; Killedar, V.; Mane, O.; Nitalikar, G.; Bhosale, S. AI-Based Medical Chatbot for Disease Prediction. *Int. J. Future Med. Res.* **2024**, *9*, 763–765. [\[CrossRef\]](#)
2. Al Kuwaiti, A.; Nazer, K.; Al-Reedy, A.; Al-Shehri, S.; Al-Muhanna, A.; Subbarayalu, A.V.; Al Muhanna, D.; Al-Muhanna, F.A. A review of the role of artificial intelligence in healthcare. *J. Pers. Med.* **2023**, *13*, 951. [\[CrossRef\]](#)
3. Kundu, R.; Shi, X.; Morrison, J.; Barrett, J.; Mukherjee, B. A framework for understanding selection bias in real-world healthcare data. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2024**, *187*, 606–635. [\[CrossRef\]](#)
4. Gonzales, A.; Guruswamy, G.; Smith, S.R. Synthetic data in health care: A narrative review. *PLoS Digit. Health* **2023**, *2*, e0000082. [\[CrossRef\]](#)
5. Mohamad, I.; Glaun, M.D.; Prabhaskar, K.; Busheri, A.; Lai, S.Y.; Noronha, V.; Hosni, A. Current treatment strategies and risk stratification for oral carcinoma. *Am. Soc. Clin. Oncol. Educ. Book* **2023**, *43*, e389810. [\[CrossRef\]](#)
6. Pierfelice, T.V.; D'Amico, E.; Cinquini, C.; Iezzi, G.; D'Arcangelo, C.; D'Ercole, S.; Petrini, M. The Diagnostic Potential of Non-Invasive Tools for Oral Cancer and Precancer: A Systematic Review. *Diagnostics* **2024**, *14*, 2033. [\[CrossRef\]](#)
7. Haj-Hosseini, N.; Lindblad, J.; Hasséus, B.; Kumar, V.V.; Subramaniam, N.; Hirsch, J.M. Early detection of oral potentially malignant disorders: A review on prospective screening methods with regard to global challenges. *J. Maxillofac. Oral Surg.* **2024**, *23*, 23–32. [\[CrossRef\]](#)
8. Dehbozorgi, R.; Zangeneh, S.; Khooshab, E.; Nia, D.H.; Hanif, H.R.; Samian, P.; Yousefi, M.; Hashemi, F.H.; Vakili, M.; Jamal-imoghadam, N.; et al. The application of artificial intelligence in the field of mental health: A systematic review. *BMC Psychiatry* **2025**, *25*, 132. [\[CrossRef\]](#)
9. Balcombe, L. AI chatbots in digital mental health. *Informatics* **2023**, *10*, 82. [\[CrossRef\]](#)
10. Saeidnia, H.R.; Firuzpour, F.; Kozak, M.; Majd, H.S. Advancing cancer diagnosis and treatment: Integrating image analysis and AI algorithms for enhanced clinical practice. *Artif. Intell. Rev.* **2025**, *58*, 105. [\[CrossRef\]](#)
11. Casheekar, A.; Lahiri, A.; Rath, K.; Prabhakar, K.S.; Srinivasan, K. A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions. *Comput. Sci. Rev.* **2024**, *52*, 100632. [\[CrossRef\]](#)
12. O'Brien, H.L.; Morton, E.; Kampen, A.; Barnes, S.J.; Michalak, E.E. Beyond clicks and downloads: A call for a more comprehensive approach to measuring mobile-health app engagement. *BJPsych Open* **2020**, *6*, e86. [\[CrossRef\]](#)
13. Li, M.; Jiang, Y.; Zhang, Y.; Zhu, H. Medical image analysis using deep learning algorithms. *Front. Public Health* **2023**, *11*, 1273253. [\[CrossRef\]](#)
14. Minutolo, A.; Damiano, E.; De Pietro, G.; Fujita, H.; Esposito, M. A conversational agent for querying Italian Patient Information Leaflets and improving health literacy. *Comput. Biol. Med.* **2022**, *141*, 105004. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Khurana, D.; Koli, A.; Khatker, K.; Singh, S. Natural language processing: State of the art, current trends and challenges. *Multimed. Tools Appl.* **2023**, *82*, 3713–3744. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Aggarwal, A.; Tam, C.C.; Wu, D.; Li, X.; Qiao, S. Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. *J. Med. Internet Res.* **2023**, *25*, e40789. [\[CrossRef\]](#)
18. Bansal, S. Oral Cancer (Lips and Tongue) Images. 2020. Available online: <https://www.kaggle.com/shivam17299/oral-cancer-lips-and-tongue-images> (accessed on 8 July 2024).
19. Vijay, S. Oral Cancer Data Dataset. 2021. Available online: <https://universe.roboflow.com/sagari-vijay/oral-cancer-data> (accessed on 8 July 2024).
20. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.; et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1248–1258. [\[CrossRef\]](#)
21. Kurniawan, M.H.; Handiyani, H.; Nuraini, T.; Hariyati, R.T.S.; Sutrisno, S. A systematic review of artificial intelligence-powered (AI-powered) chatbot intervention for managing chronic illness. *Ann. Med.* **2024**, *56*, 2302980. [\[CrossRef\]](#)
22. Bhirud, N.; Tataale, S.; Randive, S.; Nahar, S. A literature review on chatbots in healthcare domain. *Int. J. Sci. Technol. Res.* **2019**, *8*, 225–231.
23. Caldarini, G.; Jaf, S.; McGarry, K. A literature survey of recent advances in chatbots. *Information* **2022**, *13*, 41. [\[CrossRef\]](#)
24. Zhang, Z.; Zhang, L.; Yuan, X.; Zhang, A.; Xu, M.; Qian, F. A first look at gpt apps: Landscape and vulnerability. *arXiv* **2024**, arXiv:2402.15105. [\[CrossRef\]](#)
25. Babaei, N.; Zamanzadeh, V.; Valizadeh, L.; Lotfi, M.; Kousha, A.; Samad-Soltani, T.; Avazeh, M. Virtual care in the health care system: A concept analysis. *Scand. J. Caring Sci.* **2024**, *38*, 35–46. [\[CrossRef\]](#) [\[PubMed\]](#)

26. Bibault, J.E.; Chaix, B.; Nectoux, P.; Pienkowski, A.; Guillemasé, A.; Brouard, B. Healthcare ex Machina: Are conversational agents ready for prime time in oncology? *Clin. Transl. Radiat. Oncol.* **2019**, *16*, 55–59. [\[CrossRef\]](#)
27. Mowbray, T. A Survey of Deep Learning Architectures in Modern Machine Learning Systems: From CNNs to Transformers. *J. Comput. Technol. Softw.* **2025**, *4*.
28. Malaka, S. The Effect of Chatbots on the Security of Metadata. *Int. J. Theory Organ. Pract. (IJTOP)* **2024**, *4*, 75–82.
29. Sanchez, D.T.; Sartagoda, R.S. Security Threats and Security Testing for Chatbots. In *Conversational Artificial Intelligence*; Wiley: Hoboken, NJ, USA, 2024; pp. 303–318.
30. Laymouna, M.; Ma, Y.; Lessard, D.; Schuster, T.; Engler, K.; Lebouché, B. Roles, users, benefits, and limitations of chatbots in health care: Rapid review. *J. Med. Internet Res.* **2024**, *26*, e56930. [\[CrossRef\]](#)
31. Warin, K.; Limprasert, W.; Suebnukarn, S.; Jinaporntham, S.; Jantana, P.; Vicharueang, S. AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. *PLoS ONE* **2022**, *17*, e0273508. [\[CrossRef\]](#)
32. Xu, L.; Sanders, L.; Li, K.; Chow, J.C.L. Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review. *JMIR Cancer* **2021**, *7*, e27850. [\[CrossRef\]](#)
33. Olechowski, A.L.; Eppinger, S.D.; Joglekar, N.; Tomaszek, K. Technology readiness levels: Shortcomings and improvement opportunities. *Syst. Eng.* **2020**, *23*, 395–408. [\[CrossRef\]](#)
34. Voigt, P.; Von dem Bussche, A. *The EU General Data Protection Regulation (gdpr). A Practical Guide*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2017.
35. Gupta, A.; Goswami, P.; Chaudhary, N.; Bansal, R. Deploying an application using google cloud platform. In Proceedings of the 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 5–7 March 2020; pp. 236–239.
36. Atienza, R. *Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, Deep RL, Unsupervised Learning, Object Detection and Segmentation, and More*; Packt Publishing Ltd.: Birmingham, UK, 2020.
37. Jung, H.; Oh, Y. Towards better explanations of class activation mapping. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 1336–1344.
38. Hou, Y.; Wu, Z.; Cai, X.; Zhu, T. The application of improved densenet algorithm in accurate image recognition. *Sci. Rep.* **2024**, *14*, 8645. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Sharma, S.; Guleria, K. Deep learning models for image classification: Comparison and applications. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 28–29 April 2022; pp. 1733–1738.
40. Raza, S.M.; Raza, A.; Babeker, M.I.A.; Haq, Z.U.; Islam, M.A.; Li, S. Efficient citrus fruit image classification via a hybrid hierarchical CNN and transfer learning framework. *J. Food Meas. Charact.* **2025**, *19*, 356–377. [\[CrossRef\]](#)
41. Olabanji, S.O.; Oladoyinbo, O.B.; Asonze, C.U.; Oladoyinbo, T.O.; Ajayi, S.A.; Olaniyi, O.O. Effect of adopting AI to explore big data on personally identifiable information (PII) for financial and economic data transformation. *SSRN* **2024**. [\[CrossRef\]](#)
42. Zhang, Y.; Lau, R.Y.; David Xu, J.; Rao, Y.; Li, Y. Business chatbots with deep learning technologies: State-of-the-art, taxonomies, and future research directions. *Artif. Intell. Rev.* **2024**, *57*, 113. [\[CrossRef\]](#)
43. Sadasivan, C.; Cruz, C.; Dolgoy, N.; Hyde, A.; Campbell, S.; McNeely, M.; Stroulia, E.; Tandon, P. Examining patient engagement in chatbot development approaches for healthy lifestyle and mental wellness interventions: Scoping review. *J. Particip. Med.* **2023**, *15*, e45772. [\[CrossRef\]](#)
44. Harfield, S.; Purcell, T.; Schioldann, E.; Ward, J.; Pearson, O.; Azzopardi, P. Enablers and barriers to primary health care access for Indigenous adolescents: A systematic review and meta-aggregation of studies across Australia, Canada, New Zealand, and USA. *BMC Health Serv. Res.* **2024**, *24*, 553. [\[CrossRef\]](#)
45. Chen, S.; Kann, B.H.; Foote, M.B.; Aerts, H.J.; Savova, G.K.; Mak, R.H.; Bitterman, D.S. Use of artificial intelligence chatbots for cancer treatment information. *JAMA Oncol.* **2023**, *9*, 1459–1462. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.