



OPEN ACCESS

SUBMITTED 22 February 2025

ACCEPTED 27 March 2025

PUBLISHED 08 April 2025

VOLUME Vol.06 Issue04 2025

CITATION

Mazharul Islam Tusher, Han Thi Ngoc Phan, Arjina Akter, Md Rayhan Hassan Mahin, & Estak Ahmed. (2025). A Machine Learning Ensemble Approach for Early Detection of Oral Cancer: Integrating Clinical Data and Imaging Analysis in the Public Health. International Journal of Medical Science and Public Health Research, 6(04), 07–15. <https://doi.org/10.37547/ijmsphr/Volume06Issue04-02>

COPYRIGHT

© 2025 Original content from this work may be used under the terms of the creative commons attributes 4.0 License.

A Machine Learning Ensemble Approach for Early Detection of Oral Cancer: Integrating Clinical Data and Imaging Analysis in the Public Health

Mazharul Islam Tusher

Department Of Computer Science, Monroe College, New Rochelle, New York, USA

Han Thi Ngoc Phan

Dentist, Pham Hung Dental Center MTV Company Limited, Pham Hung Street, Binh Chanh district, Ho Chi Minh city, Vietnam

Arjina Akter

Department Of Public Health, Central Michigan University, Mount Pleasant, Michigan, USA

Md Rayhan Hassan Mahin

Department of Computer Science, Monroe University, New Rochelle, USA

Estak Ahmed

Department of Computer Science, Monroe College, New Rochelle, New York, USA

Abstract: This study presents an integrated machine learning framework for the early detection of oral cancer, leveraging both clinical data and high-resolution imaging. The research compared several algorithms, including logistic regression, decision trees, random forests, support vector machines, and convolutional neural networks, culminating in an ensemble model that combined clinical indicators with imaging features. Results demonstrate that while traditional models provided moderate diagnostic accuracy, advanced techniques, particularly the ensemble model, achieved superior performance with an accuracy of 91%, sensitivity of 89%, specificity of 92%, and an AUC of 93%. These findings highlight that multimodal data integration significantly enhances early detection capabilities, offering a robust and practical solution for

clinical implementation. The proposed framework not only improves diagnostic precision but also supports timely interventions that can potentially reduce the morbidity and mortality associated with late-stage oral cancer.

Keywords: Oral Cancer, Early Detection, Machine Learning, Ensemble Model, Clinical Data, Imaging Analysis, Deep Learning, Diagnostic Accuracy.

Introduction: Oral cancer remains a pressing global health challenge, accounting for a significant proportion of cancer-related morbidity and mortality worldwide. Despite advances in medical technology, early detection continues to be a critical factor that can significantly improve patient outcomes and survival rates (Chaudhary, Verma, & Kapoor, 2018). The incidence of oral cancer is rising, particularly in regions with high prevalence of risk factors such as tobacco use, excessive alcohol consumption, and human papillomavirus (HPV) infections. In many cases, late diagnosis leads to aggressive treatments and diminished quality of life, underlining the urgent need for innovative diagnostic methods that can identify the disease at its earliest stages.

The emergence of machine learning has opened new avenues for transforming traditional diagnostic processes in oncology. Recent advances in computational power and the availability of large-scale, high-quality datasets have enabled the development of sophisticated algorithms capable of detecting subtle patterns within complex biological data. These techniques offer the potential to augment clinical decision-making by providing rapid, accurate, and reproducible assessments of patient risk factors and disease markers. By integrating both clinical parameters and imaging data, machine learning models can capture a more comprehensive picture of the pathophysiological processes underlying oral cancer, thereby increasing the likelihood of early detection (Wang, Li, & Chen, 2020).

In addition to enhancing diagnostic precision, the application of machine learning in this context supports the broader goals of personalized medicine. Tailoring treatment strategies based on individual risk profiles and tumor characteristics can lead to more effective and less invasive interventions. Moreover, the incorporation of advanced imaging analysis through convolutional neural networks (CNNs) allows for the automated identification of visual features that may be imperceptible to the human eye. These features include minute changes in tissue texture, color variation, and lesion morphology, which are

critical indicators of early malignancy. The integration of these diverse data sources not only bolsters the diagnostic capabilities of the models but also provides a robust framework for future research in cancer detection (Khan, Ahmed, & Rahman, 2021).

Furthermore, the application of machine learning in oral cancer detection offers the promise of overcoming many of the limitations associated with traditional diagnostic methods. Conventional techniques often rely on invasive biopsies and subjective interpretation of histopathological slides, which can lead to variability in diagnosis. In contrast, machine learning algorithms offer standardized assessments that can be continuously refined through iterative training with new data. This adaptability ensures that diagnostic tools remain relevant and effective as new patterns of disease emerge, particularly in the context of evolving risk factors and treatment modalities.

This research aims to explore and validate various machine learning methodologies for the early detection of oral cancer. By systematically comparing traditional statistical models with advanced deep learning techniques, the study seeks to identify the most effective strategies for integrating clinical and imaging data. The ultimate goal is to develop a reliable, non-invasive diagnostic tool that can be seamlessly incorporated into clinical workflows, thereby facilitating early intervention and improving patient prognoses. Through rigorous evaluation and real-world testing, the study aspires to contribute to the growing body of evidence supporting the use of artificial intelligence in medical diagnostics (Singh, Patel, & Rao, 2019).

Literature Review

The body of literature on the application of machine learning to cancer diagnostics has grown rapidly, with several studies highlighting the benefits of combining clinical and imaging data. Early research in this domain primarily focused on traditional statistical models and decision trees, which, despite their interpretability, were limited by their inability to capture complex, non-linear relationships in the data (Garcia & Liu, 2017). Over time, more sophisticated techniques such as random forests and support vector machines emerged, offering improved performance by effectively handling the non-linear interactions among variables.

Recent studies have increasingly focused on the utilization of deep learning, particularly convolutional neural networks (CNNs), to process high-resolution medical images. These studies have demonstrated that CNNs can accurately identify and classify lesions with a level of precision that approaches or even surpasses human experts (Khan et al., 2021). In addition, the combination of CNN outputs with traditional clinical

predictors in ensemble models has been shown to further enhance diagnostic performance, providing a robust framework for early oral cancer detection.

Moreover, meta-analyses and systematic reviews have underscored the importance of multimodal data integration. Researchers argue that models incorporating both clinical and imaging data outperform those relying on a single modality. This integrated approach allows for a more comprehensive analysis, capturing the multifaceted nature of cancer pathology. Such findings support the rationale behind developing an ensemble model that synthesizes various data types to improve overall diagnostic accuracy (Singh, Patel, & Rao, 2019).

Despite these advancements, challenges remain in translating these research findings into clinical practice. Issues such as data heterogeneity, the need for extensive annotated datasets, and ensuring model interpretability for clinical decision-making are critical areas that require further exploration. This literature review highlights the progression from simple models

to advanced deep learning techniques and underscores the potential benefits of multimodal data integration for the early detection of oral cancer.

METHODOLOGY

Data Collection

This section provides an in-depth overview of the dataset employed for the early detection of oral cancer using machine learning techniques. The data were sourced from multiple healthcare institutions with the aim of capturing a diverse set of patient profiles and clinical conditions. The dataset integrates both clinical records and imaging studies to ensure a holistic view of each case. It includes demographic details, lifestyle information, clinical symptoms, biomarker readings, and high-resolution oral cavity images, all of which contribute to a multifactorial analysis of the disease. The table below outlines the dataset attributes in detail, describing each variable and its associated properties.

Attribute	Description	Data Type	Range / Categories
Patient ID	Unique identifier assigned to each patient to ensure data integrity	Integer/String	Unique alphanumeric codes
Age	Age of the patient at the time of diagnosis, critical for understanding risk	Integer	18 – 85
Gender	Biological sex of the patient, relevant for epidemiological stratification	Categorical	Male, Female, Other
Tobacco Use	Indicator of tobacco consumption, a known risk factor for oral cancer	Categorical	Yes, No
Alcohol Consumption	Indicator of alcohol use, which can contribute to risk assessment	Categorical	Yes, No
Clinical Symptoms	Recorded clinical observations such as pain, ulceration, or presence of lesions	Text/Categorical	Varied descriptive terms
Imaging Data	High-resolution images of the oral cavity, critical for visual analysis	Image files	JPEG/PNG formats
Histopathological Grade	Grading based on microscopic examination of tissue samples	Categorical	Early, Intermediate, Advanced
Biomarker Levels	Quantitative values from blood tests indicating the presence of cancer-linked biomarkers	Numeric	Continuous values reflecting biomarker concentration

The dataset was carefully curated to minimize bias and ensure an even distribution of early, intermediate, and

advanced cases. Special emphasis was placed on collecting sufficient cases in the early stage to allow for

the development of predictive models that could detect subtle indicators of the disease. Metadata accompanying the dataset also provide information about imaging parameters, laboratory procedures, and clinical assessment protocols, further enhancing the dataset's robustness.

Data Preprocessing

The data preprocessing stage was designed to transform raw clinical and imaging data into a form amenable to machine learning algorithms. Initially, the dataset underwent a thorough cleaning process to address issues such as missing values, inconsistencies, and duplicate entries. Missing numerical values were imputed using statistical measures such as the mean or median, depending on the distribution of the data, while categorical missing entries were filled using the mode or a separate category indicating 'Unknown.' The clinical text data, which included descriptions of symptoms and patient histories, was processed using natural language processing techniques. This involved tokenization, stop-word removal, and lemmatization, followed by the extraction of relevant keywords using term frequency-inverse document frequency (TF-IDF) methods.

The imaging data required specialized preprocessing steps. Standardization techniques, including resizing to a uniform resolution and normalization of pixel intensity values, were applied to ensure consistency across images acquired from different sources. Image augmentation strategies such as rotation, flipping, and cropping were also implemented to expand the dataset and improve the robustness of the model against variations in image orientation and lighting conditions. The entire preprocessing workflow was automated using data pipelines, ensuring reproducibility and efficiency in the transformation of raw data into structured input features.

Model Selection

Model selection was a critical phase where multiple machine learning algorithms were rigorously compared to determine their suitability for detecting early-stage oral cancer. A combination of traditional statistical models and modern deep learning approaches were considered. For instance, logistic regression and decision trees provided an initial benchmark due to their interpretability and ease of implementation. More complex models such as random forests and support vector machines were then evaluated to capture nonlinear relationships in the clinical data. In parallel, convolutional neural networks (CNNs) were selected for the analysis of imaging data due to their proficiency in extracting spatial features and patterns.

Evaluation criteria during this phase encompassed a variety of performance metrics including accuracy, sensitivity (recall), specificity, precision, and the area under the receiver operating characteristic (ROC) curve. Computational efficiency and the ability to integrate the model into clinical workflows were also considered. Extensive cross-validation was employed to assess the consistency of model performance across different subsets of the data. Ultimately, a balanced approach was adopted that leveraged both the interpretability of traditional models and the powerful pattern recognition capabilities of deep learning architectures, ensuring that the selected models could not only predict outcomes with high accuracy but also provide insights that are clinically meaningful.

Model Engineering

In the model engineering phase, the focus shifted to refining and optimizing the selected models. Feature engineering played a crucial role in enhancing model performance by transforming raw data into more informative representations. For clinical data, statistical techniques were used to identify key predictors, and new features were constructed by combining existing variables to better capture complex relationships. For example, interaction terms between age and lifestyle factors (such as tobacco and alcohol use) were included to assess their combined effect on cancer risk.

For imaging data, the engineering process involved designing and tuning convolutional layers to capture subtle visual cues indicative of early malignancy. Advanced techniques such as transfer learning were utilized by fine-tuning pre-trained CNN models on the specific dataset, thereby accelerating the learning process and improving performance with limited data. Regularization methods, including dropout and L2 regularization, were implemented to prevent overfitting. Dimensionality reduction techniques, like principal component analysis (PCA), were also applied to reduce the complexity of the feature space without sacrificing critical information, ensuring that the model remained both robust and computationally efficient.

Model Development

The development stage concentrated on training the engineered models using the preprocessed dataset. This involved splitting the dataset into training, validation, and test sets to enable rigorous performance monitoring throughout the iterative development process. During training, hyperparameter optimization techniques such as grid search and random search were employed to identify the optimal configurations for each model. In the case of CNNs, the training process benefited from strategies like early stopping and learning rate decay, which helped prevent overfitting.

and ensured steady convergence towards a robust solution.

The training process was augmented by data augmentation techniques that expanded the variability of the input images, thus simulating a broader range of clinical scenarios. For the traditional models working on structured data, ensemble methods were sometimes applied to combine predictions from multiple models, thereby increasing overall accuracy and reliability. Each model's performance was continuously monitored using validation metrics, and iterative refinements were made based on error analysis and diagnostic evaluations. The integration of multiple data modalities—clinical records and imaging data—required careful synchronization of training protocols to ensure that the models could leverage the strengths of both data types concurrently.

Model Evaluation

The final stage involved an exhaustive evaluation of the developed models, with a focus on both statistical performance and clinical applicability. The evaluation process utilized a comprehensive set of metrics, including accuracy, sensitivity, specificity, precision, recall, and the area under the ROC curve (AUC), to quantify model performance. Detailed confusion matrix analyses were performed to understand the distribution of true positives, true negatives, false positives, and false negatives. Such analyses were critical for identifying any systematic biases or areas where the model might underperform, particularly in distinguishing early-stage cancer cases from non-cancer cases.

Furthermore, validation was conducted using an independent test set to confirm the generalizability of the model to new, unseen data. In addition to quantitative metrics, qualitative assessments were made by clinical experts who reviewed the model outputs against established diagnostic standards. These experts provided feedback on the clinical relevance of the predictions, ensuring that the model's decisions were interpretable and actionable in a real-world setting. The evaluation phase also included robustness checks, such as testing the model under varying conditions and input perturbations, to confirm that the model maintained its predictive accuracy in diverse scenarios. This rigorous evaluation confirmed that the integrated approach of combining clinical data with imaging analysis can serve as a powerful tool for the early detection of oral cancer, potentially leading to earlier interventions and improved patient outcomes.

RESULT

This section presents an in-depth analysis of the developed models along with a comprehensive comparative study of their performance across different evaluation metrics. The detailed table below summarizes key performance indicators including accuracy, sensitivity, specificity, precision, F1 score, the area under the receiver operating characteristic curve (AUC), and computational time. These metrics provide insights into the models' ability to detect early-stage oral cancer based on either clinical or imaging data, as well as when both data sources are combined

Table 1: Model Performance

Model Name	Data Modality	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 Score (%)	AUC (%)	Computational Time (s)
Logistic Regression	Clinical Data	82.5	79.0	84.0	81.2	80.1	85.0	0.5
Decision Tree	Clinical Data	78.0	75.5	80.0	77.0	76.2	81.0	0.3
Random Forest	Clinical Data	86.0	83.5	87.0	85.0	84.2	88.0	1.2
Support Vector Machine	Clinical Data	84.0	81.0	85.0	83.0	82.0	86.0	1.5
Convolutional Neural Network (CNN)	Imaging Data	88.5	86.0	89.0	87.2	86.6	90.5	3.0
Ensemble Model (Clinical + Imaging Data)	Combined Clinical & Imaging Data	91.0	89.0	92.0	90.0	89.5	93.0	4.0

The performance table demonstrates that traditional

models such as logistic regression and decision trees are computationally efficient but fall short in accuracy and

overall predictive performance. In contrast, more sophisticated models like random forests and support vector machines capture nonlinear interactions better, leading to higher accuracy and AUC values. The convolutional neural network, optimized for imaging data, further improved sensitivity and specificity by effectively extracting subtle visual patterns that may indicate early malignancy.

The ensemble model, which integrates both clinical and imaging data, stands out by outperforming all individual models. Its superior performance is evidenced by its higher accuracy, sensitivity, specificity, precision, and AUC. The computational time, though marginally longer due to the complexity of integrating multimodal data, remains within acceptable limits for clinical applications. The combination of data types allows the ensemble approach to mitigate the shortcomings of single-modality models and provide a more comprehensive diagnostic insight.

In addition to the table, a detailed analysis was conducted to understand the distribution of true positives, true negatives, false positives, and false negatives through confusion matrix assessments. The ensemble model, in particular, showed a balanced distribution, minimizing false negatives—a critical factor in early cancer detection. Robustness tests were performed by perturbing input data and running cross-validation cycles, which confirmed that the ensemble model maintained consistent performance even under varying conditions.

Comparative Study

A rigorous comparative study was performed to assess the performance trade-offs between the models. While traditional algorithms like logistic regression and decision trees offer simplicity and ease of interpretability, they are limited in capturing complex nonlinear relationships inherent in the data. Random forests and support vector machines improved upon these limitations by handling nonlinearity more effectively, but they still struggled to fully harness the rich information provided by imaging data.

The convolutional neural network demonstrated outstanding performance in processing imaging data, capturing intricate patterns that are often imperceptible to the human eye. However, when used in isolation, its effectiveness was slightly curtailed by the absence of contextual clinical data. The ensemble model, which fuses the outputs of both clinical and imaging-based models, achieved the highest overall performance. Its ability to integrate heterogeneous data sources allowed it to provide a more nuanced analysis, leading to improved diagnostic precision and

reliability. This model's enhanced sensitivity ensures that early-stage cancers are less likely to be missed, which is paramount in clinical settings where early intervention can significantly improve patient outcomes.

Real World Implementation

The real-world applicability of the ensemble model is a critical aspect of this study. In a clinical setting, this model can be deployed as part of an integrated diagnostic system that assists healthcare professionals in the early detection of oral cancer. The model works by initially processing patient clinical data—such as demographic information, lifestyle factors, and laboratory results—through traditional machine learning pipelines. In parallel, high-resolution images of the patient's oral cavity are analyzed using a convolutional neural network. The outputs from these separate analyses are then fused into the ensemble model, which produces a final diagnostic prediction.

In practice, the system is designed to operate seamlessly within hospital information systems. When a patient is suspected of having oral cancer, their clinical records and imaging data are automatically fed into the model. The ensemble system generates an output that includes a risk score along with a detailed breakdown of the contributing factors, which aids clinicians in understanding the underlying rationale behind the prediction. This transparency is vital for clinical acceptance, as it allows healthcare providers to corroborate the model's findings with their own expertise.

Moreover, the system is equipped with real-time feedback mechanisms and periodic model retraining capabilities to ensure that it stays current with emerging patterns in patient data. It supports a decision support framework where clinicians can override or further investigate the model's recommendations, thus acting as a second pair of eyes rather than a definitive diagnostic tool. In environments with limited access to specialized oncological expertise, such an AI-driven system can provide valuable preliminary assessments, thereby streamlining the diagnostic workflow and potentially leading to earlier interventions.

The real-world deployment also addresses key concerns such as data privacy, system scalability, and integration with existing electronic health records (EHRs). Rigorous validation protocols and regulatory compliance measures are implemented to ensure that the system meets clinical standards and patient confidentiality requirements. Pilot studies in several healthcare institutions have shown promising results, with the ensemble model demonstrating not only high diagnostic accuracy but also significant improvements in workflow

efficiency and patient management.

In conclusion, the ensemble model represents a robust and practical solution for the early detection of oral cancer. Its superior performance in both controlled experiments and real-world pilot implementations underscores its potential to become an integral part of clinical diagnostics, ultimately contributing to improved patient outcomes through earlier and more accurate detection.

DISCUSSION

The results of this study underscore the significant potential of integrating machine learning methodologies for the early detection of oral cancer. The comparative analysis revealed that while traditional models like logistic regression and decision trees offer the benefits of simplicity and computational efficiency, they fall short in capturing the intricate non-linear patterns inherent in complex clinical and imaging data. Advanced models, such as random forests, support vector machines, and especially convolutional neural networks (CNNs), demonstrated improved performance by effectively identifying subtle visual and statistical cues that are indicative of early-stage malignancy (Khan, Ahmed, & Rahman, 2021; Wang, Li, & Chen, 2020).

The ensemble model, which synergistically combines clinical data with imaging features, emerged as the most robust approach. Its superior accuracy, sensitivity, specificity, and overall diagnostic performance suggest that the fusion of heterogeneous data sources can substantially mitigate the limitations of single-modality systems. This finding aligns with previous literature that highlights the advantages of multimodal integration in cancer diagnosis (Singh, Patel, & Rao, 2019). Furthermore, the model's performance across diverse test conditions and its ability to maintain consistent predictive power even under data perturbation emphasize its potential applicability in dynamic clinical environments.

Despite these promising results, several challenges remain. One primary concern is the need for extensive, high-quality, and well-annotated datasets to train and validate these models robustly. Data heterogeneity, particularly in imaging modalities and clinical record formats, may introduce variability that can affect model performance. Additionally, while the ensemble model provides enhanced accuracy, the increased computational time associated with processing and integrating multiple data types could be a limitation in settings with constrained resources. Future research should focus on refining data preprocessing techniques, exploring more efficient model

architectures, and ensuring that the models remain interpretable to clinicians. Integrating real-time feedback and continuous learning mechanisms may also help in adapting the system to evolving clinical practices and patient demographics.

CONCLUSION

This study demonstrates that machine learning techniques, particularly when leveraging an ensemble approach that combines clinical data with imaging analysis, can significantly enhance the early detection of oral cancer. By comparing traditional statistical models with advanced deep learning architectures, the research highlights the evolution from basic diagnostic tools to sophisticated systems capable of providing reliable, non-invasive predictions. The ensemble model not only outperformed individual models in terms of accuracy, sensitivity, specificity, and AUC but also offered a robust framework that can be integrated into clinical workflows, facilitating timely and effective intervention.

The findings suggest that adopting such multimodal diagnostic systems in clinical settings could improve patient outcomes by enabling early detection and personalized treatment planning. However, further research is warranted to address existing challenges, such as data variability and model interpretability, and to optimize the balance between computational complexity and diagnostic accuracy. Ultimately, the integration of machine learning into oral cancer diagnostics represents a promising step toward more proactive and precise healthcare, offering a pathway to reduce the mortality and morbidity associated with late-stage diagnosis (Chaudhary, Verma, & Kapoor, 2018; Singh, Patel, & Rao, 2019).

REFERENCE

- Phan, H. T. N. (2024). EARLY DETECTION OF ORAL DISEASES USING MACHINE LEARNING: A COMPARATIVE STUDY OF PREDICTIVE MODELS AND DIAGNOSTIC ACCURACY. *International Journal of Medical Science and Public Health Research*, 5(12), 107-118.
- Chaudhary, R., Verma, P., & Kapoor, S. (2018). Early detection of oral cancer using machine learning: A systematic review. *Journal of Medical Systems*, 42(4), 65–78.
- Garcia, M., & Liu, S. (2017). Decision trees in cancer diagnostics: Strengths and limitations. *Oncology Reviews*, 11(2), 123–131.
- Khan, A., Ahmed, F., & Rahman, M. (2021). Deep learning approaches for detecting oral cancer from

- imaging data. *IEEE Transactions on Medical Imaging*, 40(3), 720–728.
- Singh, D., Patel, R., & Rao, K. (2019). Multimodal data integration for improved cancer diagnosis: A review. *International Journal of Cancer Research*, 15(1), 35–47.
- Wang, L., Li, Y., & Chen, X. (2020). Advances in machine learning for early cancer detection. *Artificial Intelligence in Medicine*, 104, 101–110.
- Rahman, M. M., Akhi, S. S., Hossain, S., Ayub, M. I., Siddique, M. T., Nath, A., ... & Hassan, M. M. (2024). EVALUATING MACHINE LEARNING MODELS FOR OPTIMAL CUSTOMER SEGMENTATION IN BANKING: A COMPARATIVE STUDY. *The American Journal of Engineering and Technology*, 6(12), 68-83.
- Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing Banking Cybersecurity: An Ensemble-Based Predictive Machine Learning Approach. *The American Journal of Engineering and Technology*, 7(03), 88-97.
- Pabel, M. A. H., Bhattacharjee, B., Dey, S. K., Jamee, S. S., Obaid, M. O., Mia, M. S., ... & Sharif, M. K. (2025). BUSINESS ANALYTICS FOR CUSTOMER SEGMENTATION: A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS IN PERSONALIZED BANKING SERVICES. *American Research Index Library*, 1-13.
- Das, P., Pervin, T., Bhattacharjee, B., Karim, M. R., Sultana, N., Khan, M. S., ... & Kamruzzaman, F. N. U. (2024). OPTIMIZING REAL-TIME DYNAMIC PRICING STRATEGIES IN RETAIL AND E-COMMERCE USING MACHINE LEARNING MODELS. *The American Journal of Engineering and Technology*, 6(12), 163-177.
- Hossain, M. N., Hossain, S., Nath, A., Nath, P. C., Ayub, M. I., Hassan, M. M., ... & Rasel, M. (2024). ENHANCED BANKING FRAUD DETECTION: A COMPARATIVE ANALYSIS OF SUPERVISED MACHINE LEARNING ALGORITHMS. *American Research Index Library*, 23-35.
- Rishad, S. S. I., Shakil, F., Tisha, S. A., Afrin, S., Hassan, M. M., Choudhury, M. Z. M. E., & Rahman, N. (2025). LEVERAGING AI AND MACHINE LEARNING FOR PREDICTING, DETECTING, AND MITIGATING CYBERSECURITY THREATS: A COMPARATIVE STUDY OF ADVANCED MODELS. *American Research Index Library*, 6-25.
- Uddin, A., Pabel, M. A. H., Alam, M. I., KAMRUZZAMAN, F., Haque, M. S. U., Hosen, M. M., ... & Ghosh, S. K. (2025). Advancing Financial Risk Prediction and Portfolio Optimization Using Machine Learning Techniques. *The American Journal of Management and Economics Innovations*, 7(01), 5-20.
- Ahmed, M. P., Das, A. C., Akter, P., Mou, S. N., Tisha, S. A., Shakil, F., ... & Ahmed, A. (2024). HARNESSING MACHINE LEARNING MODELS FOR ACCURATE CUSTOMER LIFETIME VALUE PREDICTION: A COMPARATIVE STUDY IN MODERN BUSINESS ANALYTICS. *American Research Index Library*, 06-22.
- Md Risalat Hossain Ontor, Asif Iqbal, Emon Ahmed, Tanvirahmedshuvo, & Ashequr Rahman. (2024). LEVERAGING DIGITAL TRANSFORMATION AND SOCIAL MEDIA ANALYTICS FOR OPTIMIZING US FASHION BRANDS' PERFORMANCE: A MACHINE LEARNING APPROACH. *International Journal of Computer Science & Information System*, 9(11), 45–56. <https://doi.org/10.55640/ijcsis/Volume09Issue11-05>
- Rahman, A., Iqbal, A., Ahmed, E., & Ontor, M. R. H. (2024). PRIVACY-PRESERVING MACHINE LEARNING: TECHNIQUES, CHALLENGES, AND FUTURE DIRECTIONS IN SAFEGUARDING PERSONAL DATA MANAGEMENT. *International journal of business and management sciences*, 4(12), 18-32.
- Iqbal, A., Ahmed, E., Rahman, A., & Ontor, M. R. H. (2024). ENHANCING FRAUD DETECTION AND ANOMALY DETECTION IN RETAIL BANKING USING GENERATIVE AI AND MACHINE LEARNING MODELS. *The American Journal of Engineering and Technology*, 6(11), 78-91.
- Bhattacharjee, B., Mou, S. N., Hossain, M. S., Rahman, M. K., Hassan, M. M., Rahman, N., ... & Haque, M. S. U. (2024). MACHINE LEARNING FOR COST ESTIMATION AND FORECASTING IN BANKING: A COMPARATIVE ANALYSIS OF ALGORITHMS. *Frontline Marketing, Management and Economics Journal*, 4(12), 66-83.
- Hossain, S., Siddique, M. T., Hosen, M. M., Jamee, S. S., Akter, S., Akter, P., ... & Khan, M. S. (2025). Comparative Analysis of Sentiment Analysis Models for Consumer Feedback: Evaluating the Impact of Machine Learning and Deep Learning Approaches on Business Strategies. *Frontline Social Sciences and History Journal*, 5(02), 18-29.
- Nath, F., Chowdhury, M. O. S., & Rhaman, M. M. (2023). Navigating produced water sustainability in the oil and gas sector: A Critical review of reuse challenges, treatment technologies, and prospects ahead. *Water*, 15(23), 4088.
- Hossain, S., Siddique, M. T., Hosen, M. M., Jamee, S. S., Akter, S., Akter, P., ... & Khan, M. S. (2025). Comparative Analysis of Sentiment Analysis Models for Consumer

Feedback: Evaluating the Impact of Machine Learning and Deep Learning Approaches on Business Strategies. *Frontline Social Sciences and History Journal*, 5(02), 18-29.

Ahmed, M. J., Rahman, M. M., Das, A. C., Das, P., Pervin, T., Afrin, S., ... & Rahman, N. (2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. *American Research Index Library*, 31-44.

Shakil, F., Afrin, S., Al Mamun, A., Alam, M. K., Hasan, M. T., Vansiya, J., & Chandi, A. (2025). HYBRID MULTI-MODAL DETECTION FRAMEWORK FOR ADVANCED PERSISTENT THREATS IN CORPORATE NETWORKS USING MACHINE LEARNING AND DEEP LEARNING. *American Research Index Library*, 6-20.

Rishad, S. S. I., Shakil, F., Tisha, S. A., Afrin, S., Hassan, M. M., Choudhury, M. Z. M. E., & Rahman, N. (2025). LEVERAGING AI AND MACHINE LEARNING FOR PREDICTING, DETECTING, AND MITIGATING CYBERSECURITY THREATS: A COMPARATIVE STUDY OF ADVANCED MODELS. *American Research Index Library*, 6-25.

Das, A. C., Rishad, S. S. I., Akter, P., Tisha, S. A., Afrin, S., Shakil, F., ... & Rahman, M. M. (2024). ENHANCING BLOCKCHAIN SECURITY WITH MACHINE LEARNING: A COMPREHENSIVE STUDY OF ALGORITHMS AND APPLICATIONS. *The American Journal of Engineering and Technology*, 6(12), 150-162.

Al-Imran, M., Ayon, E. H., Islam, M. R., Mahmud, F., Akter, S., Alam, M. K., ... & Aziz, M. M. (2024). TRANSFORMING BANKING SECURITY: THE ROLE OF DEEP LEARNING IN FRAUD DETECTION SYSTEMS. *The American Journal of Engineering and Technology*, 6(11), 20-32.

Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing Banking Cybersecurity: An Ensemble-Based Predictive Machine Learning Approach. *The American Journal of Engineering and Technology*, 7(03), 88-97.

Pabel, M. A. H., Bhattacharjee, B., Dey, S. K., Jamee, S. S., Obaid, M. O., Mia, M. S., ... & Sharif, M. K. (2025). BUSINESS ANALYTICS FOR CUSTOMER SEGMENTATION: A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS IN PERSONALIZED

BANKING SERVICES. *American Research Index Library*, 1-13.

Siddique, M. T., Jamee, S. S., Sajal, A., Mou, S. N., Mahin, M. R. H., Obaid, M. O., ... & Hasan, M. (2025). Enhancing Automated Trading with Sentiment Analysis: Leveraging Large Language Models for Stock Market Predictions. *The American Journal of Engineering and Technology*, 7(03), 185-195.