Machine Learning-Based Classification of Shoulder Implant X-Rays for Manufacturer Identification

Sayyed Hasanoddin[1*], Bilavath Shiva[2], Bashaboyina Sai Gagan[2], Vugge Pallavi[2], Macharla Sai Kiran[2]

[1]Associate Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal, Telangana.

*Corresponding Email: hasanoddin.sayyed@gmail.com

ABSTRACT

Recent advancements in medical imaging and machine learning have facilitated automated classification of orthopaedic implants, improving accuracy and efficiency. Shoulder implants are used in orthopaedic surgeries, and their identification is crucial for revision procedures and post-operative assessments. Studies show that over 250,000 shoulder replacements are performed annually in the U.S., with implant misidentification contributing to 30% of revision complications. Traditional manual classification by radiologists is time-consuming and prone to errors, with accuracy rates varying between 65% and 80%, depending on expertise. This study proposes a machine learning-based approach for the automated classification of shoulder implant X-rays to identify their manufacturers, reducing dependency on manual efforts. The dataset consists of labeled X-ray images of shoulder implants categorized by manufacturer, with labels representing different implant brands and models. The preprocessing pipeline involves contrast enhancement, noise reduction, and resizing to ensure consistency. To address data imbalance, Synthetic Minority Over-sampling Technique (SMOTE) is employed, ensuring equal representation across classes. The dataset is then split into 80% training and 20% testing for model evaluation. We compare an existing Support Vector Machine (SVM) classifier, which has limitations in scalability and performance on imbalanced data, with a proposed Random Forest Classifier (RFC) model. The RFC outperforms SVM by leveraging multiple decision trees, reducing overfitting, and improving classification accuracy. Experimental results demonstrate that the RFC model achieves a significantly higher accuracy, precision, and recall. The proposed method offers a robust, automated solution for implant identification, aiding radiologists and orthopaedic surgeons in streamlining patient care and implant tracking.

Keywords: Shoulder implant classification, X-ray image analysis, Support vector machine, Random Forest classifier, Data balancing.

1. INTRODUCTION

Health system efficiency has been a growing concern over the years, with global healthcare expenditures increasing at an alarming rate. In 2018, global healthcare spending reached $8.3 trillion, and by 2023, this number soared to over $12 trillion. This massive rise in spending is attributed to aging populations, the prevalence of chronic diseases, and advancements in medical technologies, which often demand more resources and higher costs. Yet, despite the increased investments, many health systems across the world continue to face inefficiencies in service delivery, patient care, and resource management. For instance, the World Health Organization (WHO) estimates that up to 20-40% of healthcare resources are wasted through inefficiencies, which could range from delays in care delivery to redundant or unnecessary procedures. The complexity of managing health data, particularly with the rise of electronic health records (EHRs), has added to the burden, as outdated systems and fragmented processes prevent seamless data integration and real-time decision-making. These inefficiencies not only drive-up costs but also impede the ability to provide timely and high-quality care, highlighting the need for more innovative and adaptive solutions, such as the integration of machine learning in healthcare systems.

2. LITERATURE SURVEY

Lyon et al. (2021) [1] explored the potential of AI-driven optimization in the healthcare diagnostic process. Their study emphasized the significant role of AI in enhancing diagnostic accuracy and reducing human error, which traditionally slows down the process. By employing machine learning algorithms, the study highlighted improvements in decision-making, leading to more timely and accurate diagnoses. However, challenges such as data integration and system interoperability remain areas for improvement. Tripathi et al. (2021) [2] examined the evolving role of big data and AI in drug discovery. The authors discussed how AI-powered tools have accelerated the drug discovery process by analyzing vast datasets and identifying potential drug candidates more efficiently than traditional methods. While AI presents numerous advantages, the study also pointed out limitations in data handling and the complexity of integrating AI systems into existing pharmaceutical frameworks. Khan et al. (2023) [3] discussed the drawbacks of AI in the healthcare sector, emphasizing challenges such as data privacy, bias in AI algorithms, and the high cost of implementing AI solutions. The paper suggested potential solutions, including more robust data governance frameworks and continuous algorithm validation to minimize bias and improve trust in AI-driven healthcare solutions.

Dileep and Gianchandani (2022) [4] focused on the use of AI in breast cancer screening and diagnosis. The study demonstrated how AI could improve early detection rates and reduce false positives in mammograms, leading to better patient outcomes. The authors also highlighted the need for ongoing training of AI models to account for new medical data and maintain high levels of diagnostic accuracy. Chandrashekar et al. (2020) [5] introduced a deep learning approach to generate contrast-enhanced CT angiograms without the need for intravenous contrast agents. The study underscored the potential of AI to reduce patient risk by minimizing exposure to contrast agents, while still maintaining high-quality imaging for diagnostic purposes. This innovation marked a significant step forward in non-invasive imaging techniques. William et al. (2018) [6] assessed the accuracy of an AI-driven algorithm for detecting atrial fibrillation using smartphone technology. The iREAD study showed promising results, with the AI model achieving high accuracy rates in identifying atrial fibrillation, thus offering a convenient and accessible method for early detection of heart conditions, especially in remote or underserved populations.

Li et al. (2020) [7] evaluated the use of AI to detect COVID-19 and community-acquired pneumonia using pulmonary CT scans. The study found that AI models could effectively distinguish between COVID-19 and other respiratory conditions, providing a valuable diagnostic tool during the pandemic. The authors also highlighted the importance of rapid AI adaptation to emerging diseases and updated data.Olive-Gadea et al. (2020) [8] developed a deep learning-based software capable of identifying large vessel occlusions on noncontrast CT scans. The study demonstrated that AI could significantly enhance the speed and accuracy of stroke diagnosis, enabling faster intervention and potentially improving patient outcomes in critical care settings. Lin et al. (2022) [9] discussed the role of AI-driven decision-making in the auxiliary diagnosis of epidemic diseases. The study focused on how AI models could analyze vast datasets in real-time to predict outbreaks and assist in the early diagnosis of diseases. By integrating machine learning techniques, healthcare systems could become more adaptive and responsive to emerging public health threats.

Iqbal et al. (2022) [10] conducted a narrative review on the future of AI in neurosurgery. The authors concluded that AI holds great promise in areas such as surgical planning, real-time decision support, and postoperative care. However, the review also pointed out the need for more extensive clinical validation of AI systems before they can be widely adopted in neurosurgical practices. Nguyen et al. (2018) [11] explored deep learning for sudden cardiac arrest detection in automated external defibrillators (AEDs). The study showed that AI models could significantly enhance AED functionality by improving the accuracy of sudden cardiac arrest detection, potentially saving lives in critical situations where rapid response is essential.Mostafa et al. (2022) [12] conducted a survey on AI techniques used for thoracic disease diagnosis via medical images. The study highlighted the advances in deep learning models that have improved diagnostic accuracy for conditions such as lung cancer and pneumonia. However, the survey also identified challenges related to the interpretability of AI models and the need for further research in this area.

Comito et al. (2022) [13] discussed the application of AI-driven clinical decision support systems in enhancing disease diagnosis by exploiting patient similarity. Their research demonstrated how machine learning could identify patterns in patient data to offer personalized diagnostic and treatment options, improving overall patient care and operational efficiency in hospitals.Brinker et al. (2019) [14] conducted a study comparing the diagnostic capabilities of deep neural networks to those of dermatologists in melanoma image classification. The results showed that AI outperformed human experts, marking a significant advancement in the use of AI for dermatological diagnostics and opening up new possibilities for remote or automated skin cancer screening.Santosh and Gaur (2021) [15] provided a comprehensive overview of AI and machine learning applications in public healthcare. They discussed how AI can address various public health issues, such as disease surveillance, outbreak prediction, and resource optimization, thereby improving health system efficiency on a large scale. The book emphasized the need for ethical AI practices and robust frameworks to ensure responsible implementation of these technologies in public health.

3. PROPOSED METHODOLOGY

The project presents a machine learning framework for shoulder implant manufacturer identification, combining advanced image preprocessing, synthetic data balancing (SMOTE), and an optimized classifier selection approach—a methodology not previously explored in existing surveys. Unlike conventional SVM-based approaches, which suffer from scalability and imbalanced datasets, this method integrates Random Forest Classifier (RFC) for improved generalization and decision-making. The pipeline begins with image preprocessing techniques such as contrast enhancement, noise filtering, and morphological operations to refine implant structures. SMOTE is applied after feature extraction to address class imbalance, ensuring fair representation of all implant categories. The dataset is then split into training (80%) and testing (20%), where SVM serves as a baseline model while RFC is proposed as a superior alternative. The experimental results demonstrate that RFC outperforms SVM in accuracy, recall, and robustness against imbalanced data, making it an effective automated solution for implant classification.

Step-1: Data Acquisition and Preprocessing

The dataset consists of shoulder implant X-ray images with labels representing different implant manufacturers. To ensure consistency and improve classification performance, data preprocessing is applied. The key steps include:

- Resizing the images to a uniform dimension to maintain consistency across the dataset.
- Standardization of pixel intensity values to ensure uniformity in brightness and contrast.

This preprocessing step prepares the dataset for effective feature extraction and classification.

Step-2: Handling Class Imbalance Using SMOTE

Since some implant categories have fewer samples than others, the dataset is imbalanced. To address this, Synthetic Minority Over-sampling Technique (SMOTE) is used to generate synthetic samples for underrepresented classes. This ensures that all implant categories have sufficient representation, preventing bias during model training.

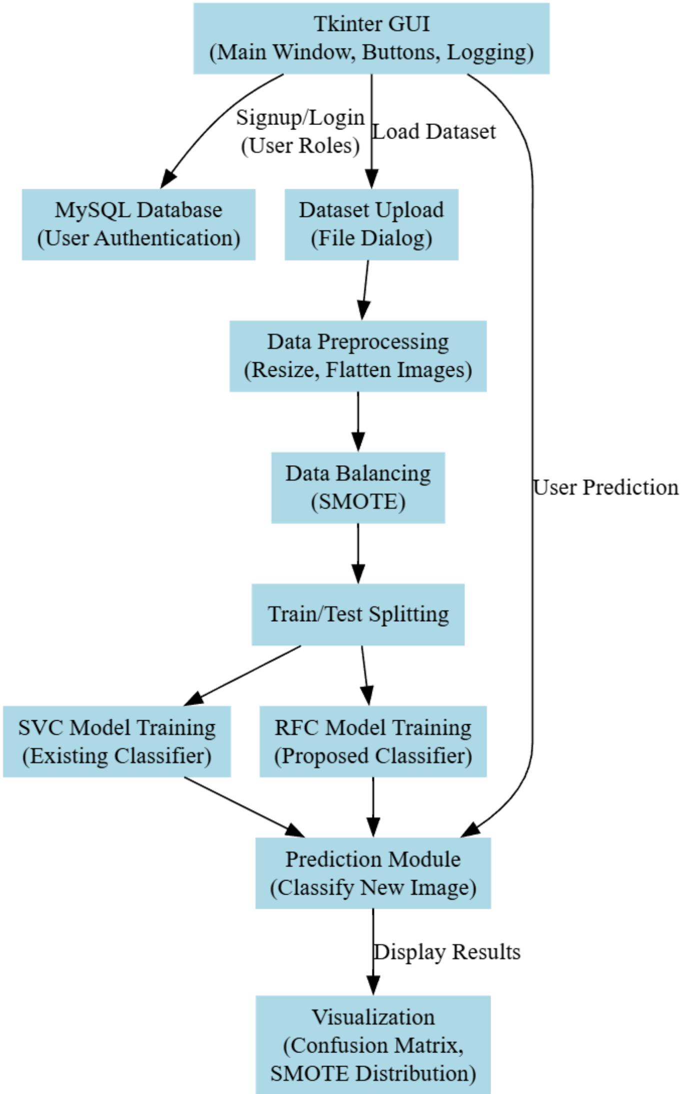


Fig. 1: Proposed system architecture of ML-based classification of shoulder implant X-rays for manufacturer identification.

Step-3: Train-Test Splitting and Baseline Model (SVM)

The dataset is then split into 80% training and 20% testing to evaluate model performance. As a baseline, an SVM (Support Vector Machine) classifier is implemented. SVM is a widely used classification algorithm, but it has limitations in handling large datasets and imbalanced distributions, which can affect classification accuracy.

Step-4: Proposed Model: Random Forest Classifier (RFC)

To improve classification performance, this study proposes the Random Forest Classifier (RFC) as an alternative to SVM. RFC is an ensemble learning algorithm that constructs multiple decision trees and aggregates their predictions. Compared to SVM, RFC provides better generalization, improved robustness to imbalanced data, and higher accuracy in identifying implant manufacturers.

Step-5: Model Evaluation

The models (SVM and RFC) are evaluated using classification metrics, including:

- **Accuracy** – Measures the overall correctness of predictions.
- **Precision & Recall** – Evaluates the reliability of classification results.
- **F1-Score** – Ensures balanced evaluation for imbalanced datasets.

3.2 ML Model Building

3.2.1 SVM

Support Vector Machine (SVM) is a supervised learning algorithm used for classification tasks. It works by finding the optimal decision boundary that maximizes the margin between different classes.The SVM classifier is implemented efficiently to classify shoulder implant manufacturers based on X-ray images.An SVM classifier is implemented to classify shoulder implant X-rays based on manufacturer labels. The process begins by checking if a pre-trained SVM model (SVM_model.pkl) exists; if found, it is loaded using joblib to avoid retraining. If not, a new Support Vector Classifier (SVC) is initialized with a polynomial kernel (poly), which transforms input data into a higher-dimensional space to handle complex decision boundaries. The regularization parameter (C=1.0) balances misclassification tolerance, while gamma='scale' determines the impact of each training sample on the decision boundary. The model is trained using X_train (input features) and y_train (manufacturer labels), learning to classify X-ray images accurately. Once trained, the model is saved for future use, ensuring efficiency. During testing, the trained SVM model predicts manufacturer labels for unseen X_test data, and the predictions (y_pred_bnb) are compared to actual labels (y_test) as shown in Fig.4.1. The function then evaluates classification performance using metrics like accuracy, precision, recall, and F1-score to measure how well the model differentiates between manufacturers. Although SVM is effective in high-dimensional spaces and for complex classification tasks, it has drawbacks such as being computationally expensive, sensitive to noise, and requiring careful kernel selection, which may limit its scalability for large datasets.

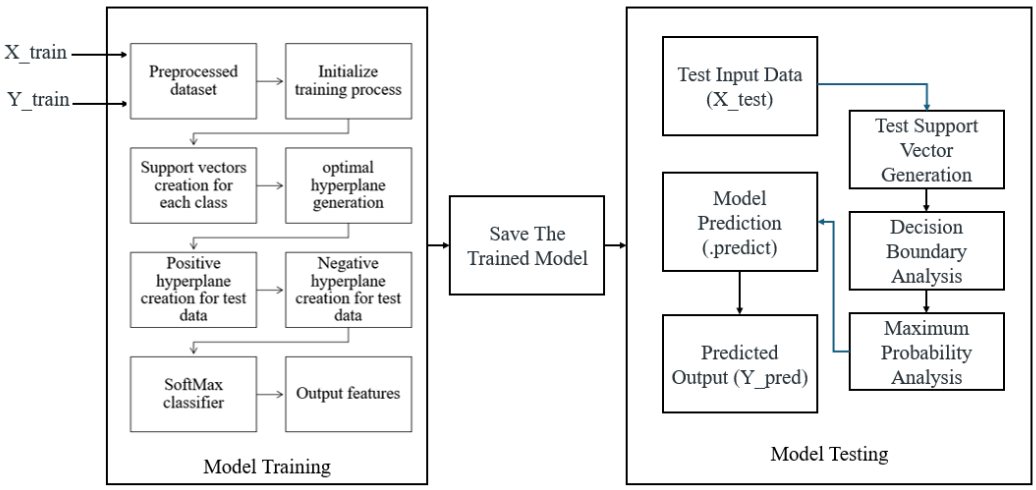


Fig. 2: Internal workflow of SVC model.

Step 1: Initializing Variables for Performance Metrics

The function first initializes four global lists: accuracy, precision, recall, and fscore. These lists will store the evaluation metrics of the Support Vector Machine (SVM) classifier after testing. These metrics help assess how well the model performs in classifying the shoulder implant X-ray images based on manufacturer labels.

Step 2: Checking for an Existing SVM Model

The function then checks if a pre-trained SVM model (SVM_model.pkl) exists in the model directory. If the file is found, it is loaded using joblib, allowing the model to be reused without retraining. This step is efficient as it saves time by avoiding unnecessary retraining.

Step 3: Training the SVM Classifier with Training Data

If the model is not found, the function initializes a new Support Vector Classifier (SVC) with the following parameters:

kernel='poly': A polynomial kernel is used to map input features into higher-dimensional space, allowing for better separation of complex data.

C=1.0: This parameter controls the trade-off between achieving a low error and maximizing the margin.

gamma='scale': Determines how much influence a single training example has, affecting the decision boundary's flexibility.

random_state=42: Ensures reproducibility by setting a fixed random seed.

The classifier is then trained using X_train and y_train, where:

X_train: Represents the input features of shoulder implant X-rays.

y_train: Represents the manufacturer labels corresponding to each X-ray.

The trained SVM model is then saved as a file (SVM_model.pkl) for future use, preventing the need for retraining unless the dataset changes.

Step 4: Predicting Labels for Test Data

Once trained, the model is used to predict the labels for test data (X_test). The predict function takes X_test (new shoulder implant X-ray images) as input and assigns each image to a predicted manufacturer category (y_pred_bnb). This output is then compared to the actual labels (y_test) to assess the model's performance.

Step 5: Evaluating the SVM Model

After prediction, the function calculates performance metrics such as accuracy, precision, recall, and F1-score. These metrics help measure how well the model correctly classifies the shoulder implant manufacturers. The results are stored in the previously initialized lists for analysis.

3.2.2 RFC Model

Random Forest Classifier (RFC) is an ensemble learning algorithm that builds multiple decision trees and combines their outputs to make accurate predictions. Each decision tree in the forest is trained on a different random subset of the dataset, allowing RFC to capture various patterns in the data as shown in Fig.1. During classification, each tree independently predicts the class label, and the final decision is determined by majority voting. This ensemble approach makes RFC more robust to overfitting, noise, and variations in data compared to single decision trees. It is particularly useful for handling high-dimensional datasets, complex feature interactions, and imbalanced data, making it an effective choice for classifying shoulder implant X-ray images based on manufacturer labels.

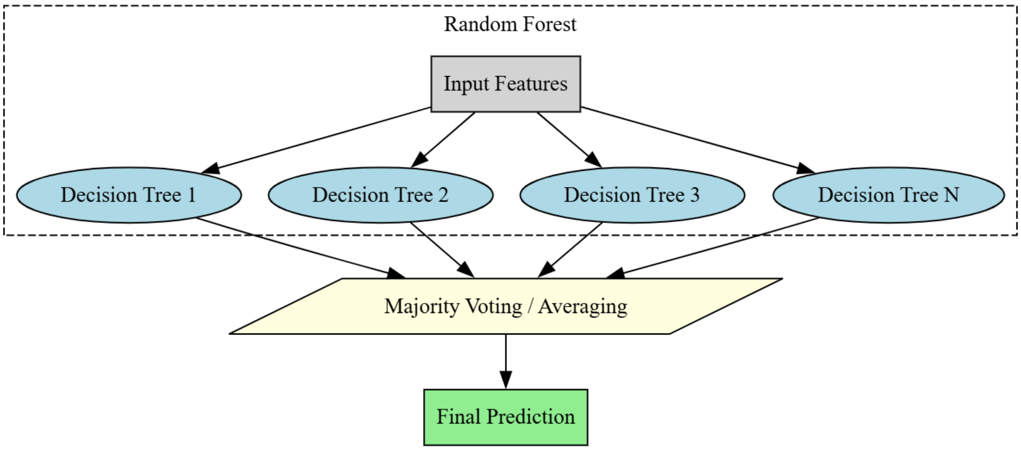


Fig. 3: RFC classifier workflow.

Step 1: Checking for an Existing Model

The function first checks if a pre-trained Random Forest Classifier (RFC) model (RFC_Model.pkl) exists in the model directory. If the file is found, it is loaded using joblib, allowing the system to reuse the trained model without retraining. This approach saves computational time and ensures efficiency when handling large datasets.

Step 2: Initializing and Training the RFC Model

If no pre-trained model is found, a new RandomForestClassifier is instantiated and trained using the dataset. The training process involves:

- X_train as input data → This contains the preprocessed X-ray images of shoulder implants in numerical format.
- y_train as output labels → These labels indicate the manufacturer category corresponding to each X-ray image.

The Random Forest algorithm works by creating multiple decision trees, each trained on random subsets of the dataset. During training, each tree learns different decision boundaries to classify the X-ray images. By aggregating predictions from multiple trees, RFC improves accuracy and reduces overfitting.

Step 3: Saving the Trained Model

Once the model is trained, it is saved using joblib.dump(), ensuring that the classifier can be used later without needing to retrain it. This step enhances efficiency, especially when working with large datasets that require significant training time.

Step 4: Making Predictions on Test Data

The trained RFC model is then tested using X_test (unseen X-ray images). The predict function is applied to classify each test sample into a manufacturer category, producing a set of predicted labels (y_pred). These predictions are compared with the actual manufacturer labels (y_test) to evaluate the model's performance.

Step 5: Evaluating the Model

The predicted labels (y_pred) are passed to the calculateMetrics function, which computes performance metrics such as accuracy, precision, recall, and F1-score. These metrics help assess the classifier's effectiveness in identifying shoulder implant manufacturers based on X-ray images. By using an ensemble of decision trees, RFC ensures robust classification performance with improved generalization compared to single classifiers like SVM.

5.RESULTS AND DISCUSSION

This research is a complete Python application that combines machine learning, image processing, and a graphical user interface (GUI) to classify shoulder implant X-rays and identify their manufacturers. Fig. 4 (left) displays the count plot, which shows the number of samples available in each category. It provides insight into the distribution of the dataset and helps identify potential class imbalances. Here, some categories have significantly fewer samples than others, which could impact model performance. Thus, SMOTE (Synthetic Minority Over-sampling Technique) is used to address such imbalances by oversampling the minority classes to ensure that the classifier is not biased toward the majority class as demonstrated in Fig. 4 (right).

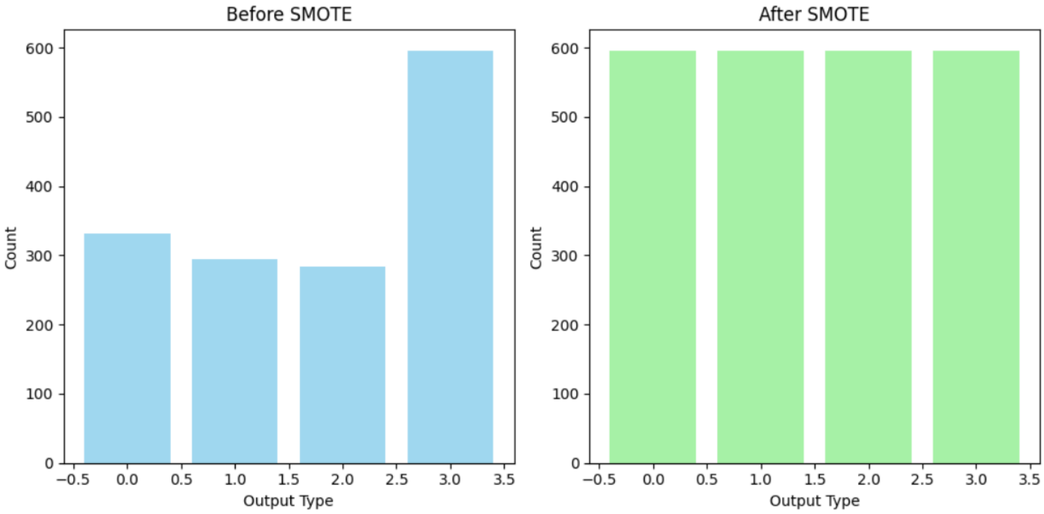
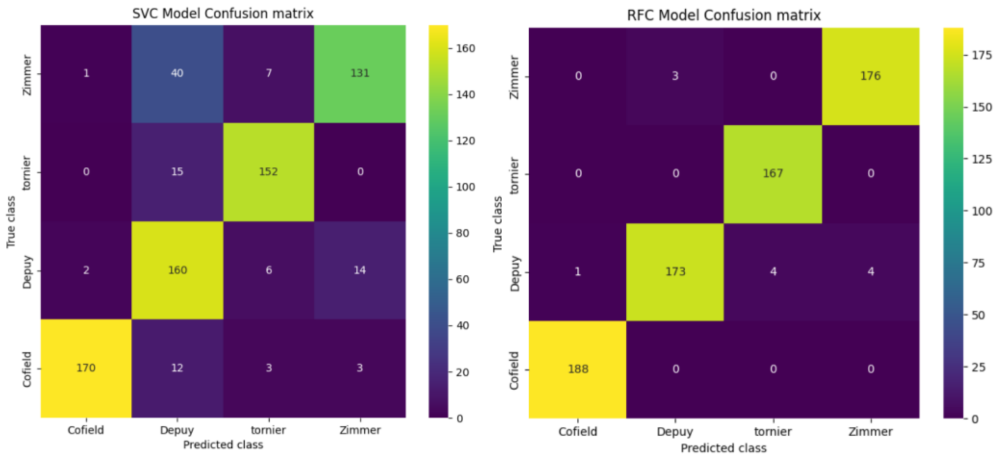


Fig. 4: Count distribution versus type of manufacturer. Before applying SMOTE (left). After applying SMOTE (right).



(a) (b)

Fig. 5: Confusion matrices obtained using (a) SVC model. (b) Proposed RFC model.

Fig. 5 presents the confusion matrices obtained using SVC model, and proposed RFC model. The confusion matrix summarizes the performance of the models by showing how many instances were correctly or incorrectly classified. Each cell represents the number of true positive, false positive, true negative, and false negative predictions. This matrix is vital for evaluating the classifier's precision, recall, and accuracy, giving a deeper understanding of how well the model is performing across all categories. This figure allows you to compare the two classifiers side-by-side and decide which model provides better results based on metrics such as accuracy, precision, and recall for each category.

Table. 1: Performance comparison of quality metrics obtained using SVC model, and proposed RFC model.

Model/Metrics	Accuracy	Precision	Recall	F1-score
SVC model	85.61	86.93	85.63	85.82
Proposed RFC model	98.32	98.30	98.34	98.31

Table. 1 presents a side-by-side comparison of key performance metrics between the SVC model and the proposed RFC model. The metrics include Accuracy, Precision, Recall, and F1-score. The SVC model records an accuracy of 85.61%, with precision, recall, and F1-score around 86.93%, 85.63%, and 85.82% respectively. In contrast, the proposed RFC model significantly outperforms the SVC, achieving an accuracy of 98.32% and nearly identical precision, recall, and F1-score values (98.30%, 98.34%, and 98.31% respectively). This comparison clearly highlights the superior performance of the RFC model in accurately classifying shoulder implant X-rays for manufacturer identification, demonstrating its robust predictive capabilities.

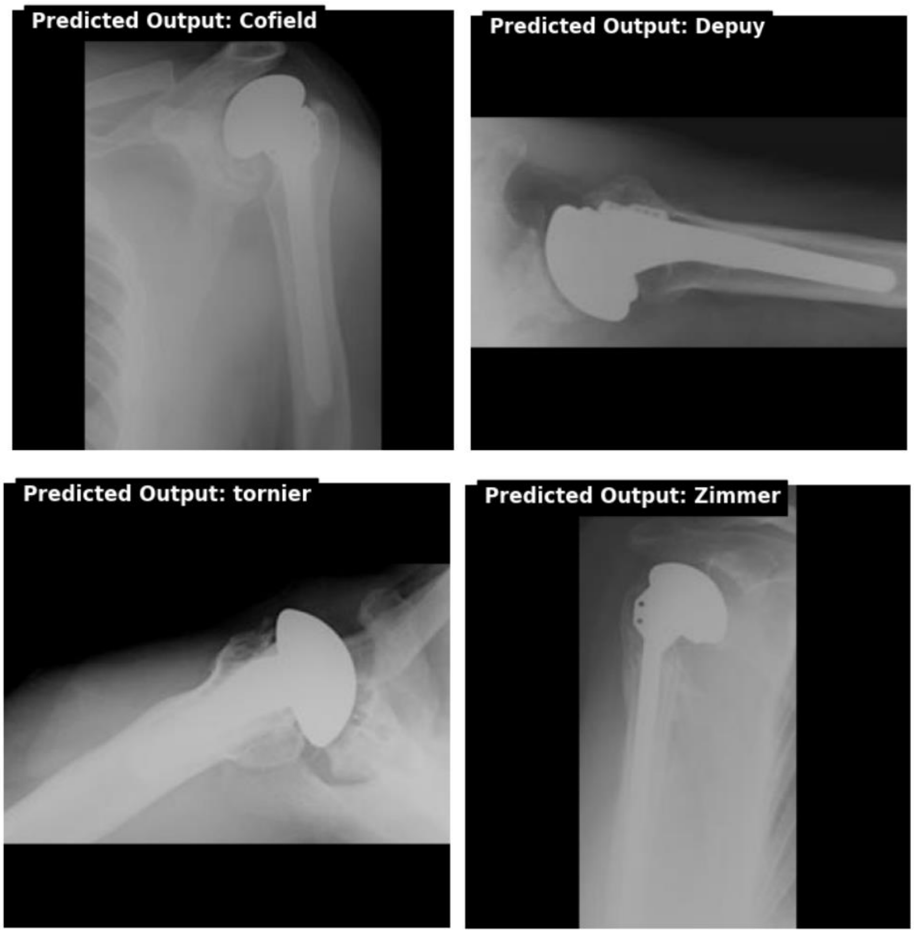


Fig. 6: Sample predictions on test images.

Fig. 6 illustrates the outcome of the prediction on sample test data using proposed RFC model. The test image is processed through the classification pipeline, and the predicted category is displayed along with the actual image. The visual overlay of the prediction on the test image provides an intuitive understanding of the model's effectiveness in real-world scenarios. It helps in verifying whether the model is correctly classifying unseen data.

5. CONCLUSION

The project demonstrates a comprehensive approach to classifying shoulder implant X-ray images for manufacturer identification using machine learning techniques. The system integrates a robust preprocessing pipeline that includes image resizing, flattening, and class balancing using SMOTE to ensure reliable input data. Two distinct classification models were developed—a Support Vector Classifier (SVC) and a Random Forest Classifier (RFC)—to compare their performance in accurately identifying the manufacturer from the processed X-ray images. While the SVC model achieved respectable performance with an accuracy of 85.61% and closely matched precision, recall, and F1-scores, the proposed RFC model significantly outperformed it, boasting an accuracy of 98.32% with similarly high values for precision, recall, and F1-score. This superior performance is primarily attributed to the ensemble nature of the Random Forest, which combines multiple decision trees and leverages majority voting to reduce overfitting and handle high-dimensional data effectively. The application is designed with an intuitive GUI using Tkinter, which supports both administrative (hospital) and user (patient) functionalities. The admin interface enables hospital staff to manage dataset uploads, preprocessing, model training, and evaluation, while the user interface allows patients to log in and test new X-ray images easily. This role-based design not only enhances security but also ensures that each user type can access the specific functionalities tailored to their needs. The integration of a MySQL database for user authentication further strengthens the system's reliability and usability.

REFERENCES

[1] AI-driven optimization in healthcare: the diagnostic process. Lyon J, Bogodistov Y, Moormann J. *Eur J Manage Issues.* 2021; 29:218–231

[2] Evolving scenario of big data and artificial intelligence (AI) in drug discovery. Tripathi MK, Nath A, Singh TP, Ethayathulla AS, Kaur P. *Mol Divers.* 2021;25:1439–1460

[3] Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. Khan B, Fatima H, Qureshi A, Kumar S, Hanan A, Hussain J, Abdullah S. *Biomed Mater Devices.* 2023:1–8.

[4] Artificial intelligence in breast cancer screening and diagnosis. Dileep G, Gianchandani Gyani SG. *Cureus.* 2022;14:0

[5] A deep learning approach to generate contrast-enhanced computerized tomography angiograms without the use of intravenous contrast agents. Chandrashekar A, Shivakumar N, Lapolla P, et al. *Eur Heart J.* 2020;41:0.

[6] Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: the iREAD Study. William AD, Kanbour M, Callahan T, et al. *Heart Rhythm.* 2018;15:1561–1565.

[7] Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Li L, Qin L, Xu Z, et al. *Radiology.*

[8] Deep learning based software to identify large vessel occlusion on noncontrast computed tomography. Olive-Gadea M, Crespo C, Granes C, et al. *Stroke*

[9] AI-driven decision making for auxiliary diagnosis of epidemic diseases. Lin K, Liu J, Gao J. *IEEE Transact Mol Biol Multi-Scale Commun.* 2022;8:9–16.

[10] The future of artificial intelligence in neurosurgery: a narrative review. Iqbal J, Jahangir K, Mashkoor Y, et al. *Surg Neurol Int.* 2022;13:536.

[11] Deep feature learning for sudden cardiac arrest detection in automated external defibrillators. Nguyen MT, Nguyen BV, Kim K. *Sci Rep.* 2018;8:17196

[12] A survey on AI techniques for thoracic diseases diagnosis using medical images. Mostafa FA, Elrefaei LA, Fouda MM, Hossam A. *Diagnostics (Basel)* 2022;12:3034.

[13] AI-driven clinical decision support: enhancing disease diagnosis exploiting patients similarity. Comito C, Falcone D, Forestiero A. *IEEE Access.* 2022;10:6878–6888.

[14] Deep neural networks are superior to dermatologists in melanoma image classification. Brinker TJ, Hekler A, Enk AH, et al. *Eur J Cancer.* 2019;119:11–17.

[15] Santosh K, Gaur L. *Artificial Intelligence and Machine Learning in Public Healthcare.* Singapore: Springer; 2021. AI solutions to public health issues; pp. 23–32.