BONE FRACTURE CLASSIFICATION

COURSE MACHINE LEARNING
MASTER DEGREE IN 'INGEGNERIA E SCIENZE INFORMATICHE'

2023/2024

NARDI ANNA
MATRICOLA VR487584

INDEX

WHY BONES FRACTURE?

The integration of machine learning techniques in the medical field in the recent years has brought an amazing contribution.
One area where these technologies can have a significant impact is radiology particularly in the identification and classification of bone fractures as these represent one of the most common injuries treated in the emergency department.

THE PROBLEM

The problem of identifying bone fractures in radiology images is critical in medical diagnostics and comes with challenges such as time consumption and potential for human error. Radiologists often face the task of examining numerous images to detect fractures accurately, which can be both labor-intensive and prone to variability in interpretation.

Challenges in Bone Fracture Identification

1. **Complexity and Variability**: Bone fractures present in various shapes, sizes, and locations, making their detection a complex task. Small or subtle fractures may be easily overlooked or misinterpreted, leading to delayed diagnosis or incorrect treatment plans.
2. **Manual Inspection**: Traditionally, radiologists rely on visual inspection and their expertise to identify fractures from X-ray or CT scan images. This process is time-consuming and can vary based on individual skill and experience levels, potentially affecting diagnostic accuracy.
3. **Subjectivity and Error**: Human interpretation of radiological images introduces subjective factors that can influence diagnostic outcomes. Factors such as fatigue, distraction, or varying levels of expertise among radiologists may impact the reliability of fracture detection.

IMPORTANCE

The correct identification and localization of bone fractures is important for treatment planning and complication prevention.
Diagnostic error can lead to wrong treatments.
In emergency departments, offering support through an automating diagnostic process can:

- Improve diagnostic accuracy.
- Decrease radiologists' workload.
- Improve efficiency, accelerating the diagnostic process.

RESEARCHS' THEMES

There is an extensive research dedicated to find the optimal machine learning algorithms for classification of bone structures, with several prominent approaches such as SVM (support vector machine), Random Forster and CNN (convolutional Neural Networks).

SVMs are widely used for their effectiveness in handling high dimensional data. They find the best hyperplane that maximizes the margin between different classes in the feature space. This algorithm has shown really good performance.

Random Forest combines multiple decision trees to make predictions. This methods can handle large dataset with high dimensionality and complex relationship between features.

CNN excel at learning hierarchical representations of features directly from raw data. They are a powerful tool for feature extraction in medical images.

The quest to find the best machine learning algorithms for bone fracture classification is an active field.

STATE OF ART

The field of bone fracture classification has seen significant advancements in recent years, driven by the application of machine learning techniques. As of 2024, several studies have contributed to establishing the current state-of-the-art in this domain.

Early Approaches and Techniques:

Studies such as "Detection and Classification of Long Bone Fractures" and "Automatic Detection of Fractures in Femur Bones" have utilized traditional machine learning methods combined with image processing techniques. These include preprocessing steps to enhance image quality, feature extraction methods like the Bag of Words model or gray-level Co-occurrence matrix, and classification algorithms such as Support Vector Machines (SVM), Random Forest, Nearest Neighbors, Decision Trees, and Naïve Bayes.

"Detection and Classification of Long Bone Fractures" (Avinash Vishnu V, 2017) employed a Bag of Words model for feature extraction and SVM for classification, achieving an accuracy rate of 78%. Meanwhile, "Automatic Detection of Fractures in Femur Bones" used SVM with an 84.7% accuracy rate.

Performance of Different Models: The study "Detection of Bone Fracture Based on Machine Learning Techniques" (Ahmed, 2023) evaluated multiple classification models:

- SVM demonstrated high accuracy at 92.8%.
- Random Forest achieved 85.7% accuracy.
- Nearest Neighbors reached 83.9% accuracy.
- Decision Tree showed an 80% accuracy.
- Naïve Bayes achieved 64.2% accuracy.

These results highlight the effectiveness of SVM for bone fracture classification due to its ability to handle complex feature spaces and generalize well.

Advancements in Deep Learning: More recent advancements leverage deep learning techniques, particularly Convolutional Neural Networks (CNNs) and their variants. CNNs have shown promise in automatically learning hierarchical features from medical images without the need for explicit feature extraction steps. This approach often surpasses traditional methods by capturing intricate patterns and structures within the images.

Challenges and Future Directions: Despite these advancements, challenges remain, such as the need for large annotated datasets, robustness to variations in imaging conditions, and interpretability of deep learning models. Future research is likely to focus on integrating multimodal data (e.g., combining X-rays with patient history), improving model interpretability, and enhancing real-time deployment in clinical settings.

While traditional machine learning methods have laid a strong foundation for bone fracture classification, the shift towards deep learning signifies a promising future for more accurate and automated diagnostic systems in orthopedics. This evolution not only improves diagnostic accuracy but also holds potential for enhancing patient care through faster and more reliable fracture detection and classification.

OBJECTIVES

1. **Automated Classification and Localization:**
   - **Goal:** Develop a robust system capable of automatically classifying bone fractures in radiology images using machine learning and deep learning techniques.
   - **Implementation:** Implement and integrate machine learning algorithms and deep learning models to accurately identify bone fractures in the provided radiology images.
2. **Model Comparison:**
   - **Goal:** Evaluate and compare the performance of various models to determine the most effective one for bone fracture classification.
   - **Implementation:** Train, test, and compare different machine learning and deep learning models, using various performance metrics to select the best-performing model.

Data Collection

1. **Dataset Compilation:**

   - **Goal:** Gather a diverse dataset of radiology images that include both fractured and non-fractured bones.
   - **Implementation:** Use publicly available datasets, store the dataset in an accessible format for processing.

2. **Data Cleaning and Preprocessing:**

   - **Goal:** Clean and process the dataset to ensure high quality and consistency.
   - **Implementation:** Implement code to avoid reading corrupted images, normalize image sizes, and standardize pixel values.

Model Development

1. **Model Selection and Implementation:**

   - **Goal:** Explore and implement various machine learning and deep learning models for fracture classification and localization.
   - **Implementation:** Write code to implement traditional machine learning models (e.g., KNN, SVM) and deep learning architectures (e.g., CNNs, VGG16). Use libraries such as TensorFlow, Keras, or PyTorch for model development.

2. **Feature Extraction and Engineering:**

   - **Goal:** Extract relevant features from radiology images to enhance model performance.
   - **Implementation:** It was utilized convolutional layers for automatic feature extraction

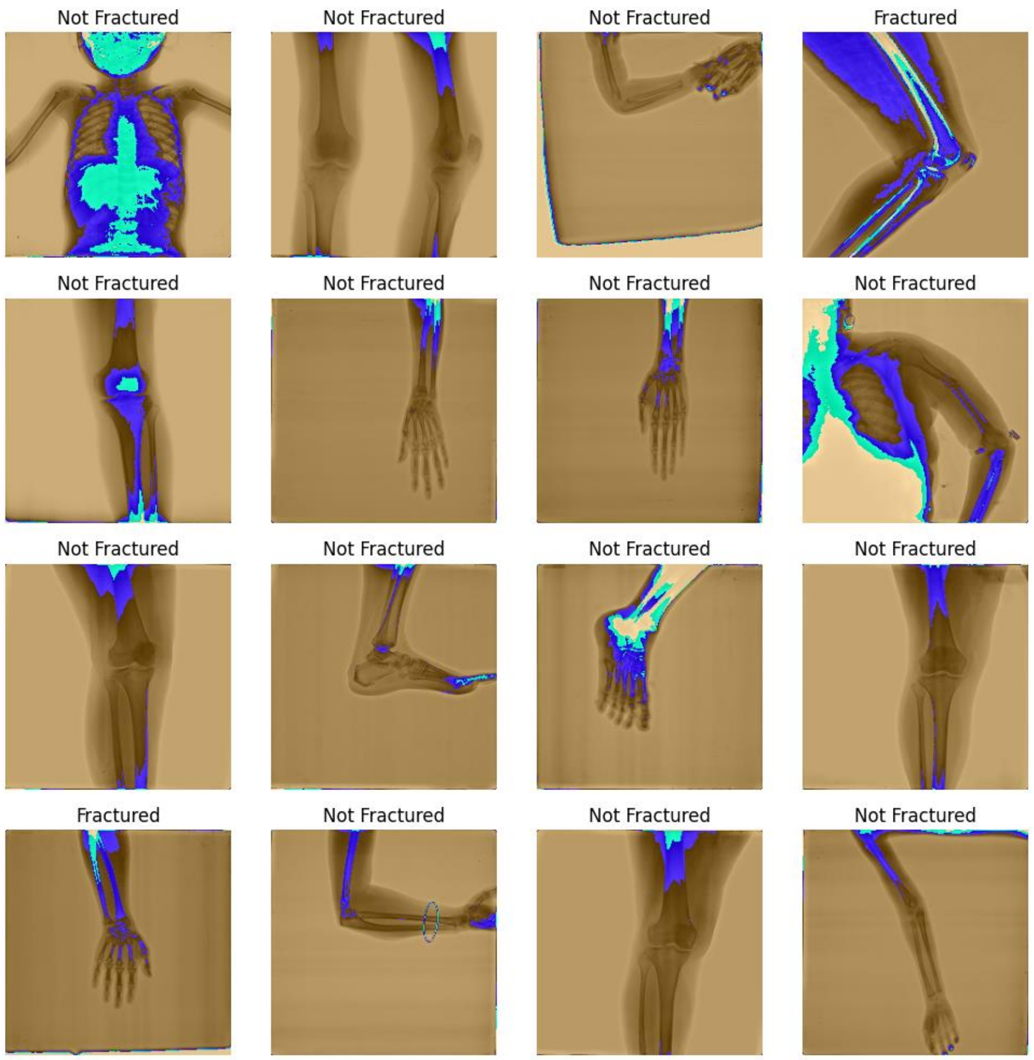
Model Evaluation

1. **Performance Metrics:**

   - **Goal:** Evaluate the models using comprehensive performance metrics.
   - **Implementation:** Implement code to calculate and display metrics such as accuracy, precision and recall. These metrics will provide a detailed view of the model's performance.

2. **Comparison and Selection:**

   - **Goal:** Compare the results of different models to identify the best-performing approach.
   - **Implementation:** Use code to generate confusion matrices visualization. Analyze these results to discuss the trade-offs and select the best model based on accuracy, and interpretability.



METHODOLOGY

Pre-processing

Image processing is a preliminary step that allows to improve image accuracy and ensure that the data is in a format suitable for the machine learning model.

Steps:

a. Converting images to RGB format: to ensure consistency and compatibility with the VGG16 model all images are converted to RGB.
b. Resizing images to 244x244 pixels to avoid issues.
c. Transforming images into Numpy array to facilitate the handling of pixel values as numerical data which is essential for model training.
d. Normalization of pixel values to improve model performance and training stability.

Feature Extraction

Feature extraction is a crucial step in a machine learning project for bone fracture classification. By flattening images, we can utilize traditional machine learning algorithms like KNN and SVM. However, these methods may not capture the complex patterns and structures within the images. To address this, we leverage the convolutional base of deep learning models like VGG16 for feature extraction.

CNNs, such as VGG16, are specifically designed to handle image data. They excel at capturing spatial hierarchies and patterns through layers of convolutions. This hierarchical feature extraction allows CNNs to detect low-level features like edges and textures in the initial layers, and more complex features like shapes and objects in deeper layers.

By using the pre-trained VGG16 model, which has been trained on a large and diverse dataset like ImageNet, we benefit from its ability to recognize a wide range of features that are relevant to image classification tasks. This transfer learning approach enables us to utilize the powerful feature extraction capabilities of VGG16 without needing to train a deep network from scratch, which is particularly beneficial when working with limited data.

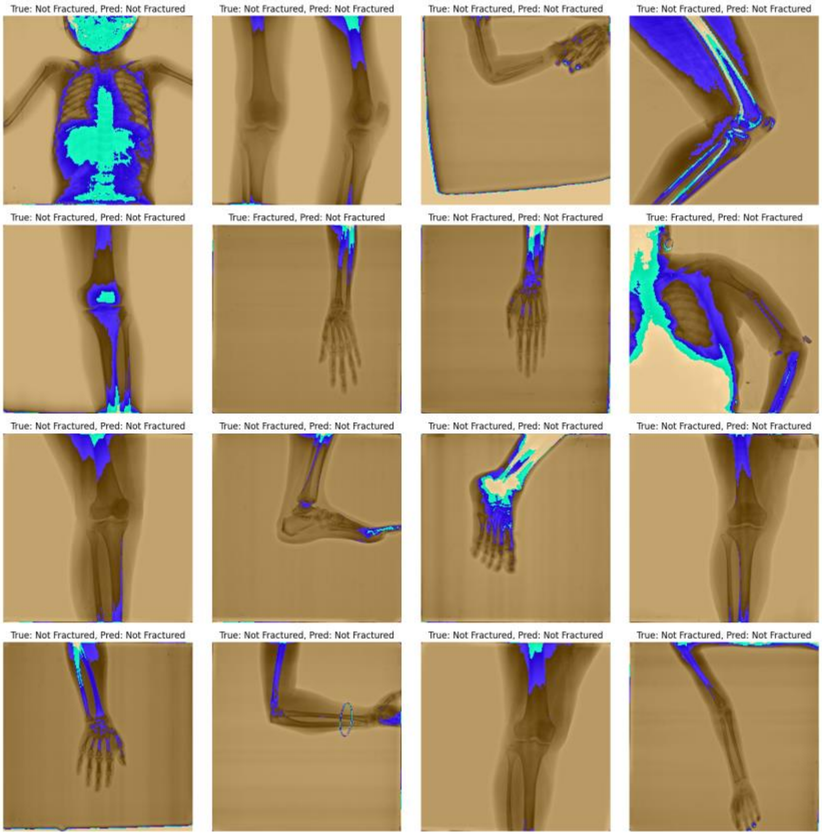
Moreover, the features extracted by CNNs are often more informative compared to traditional methods, leading to improved performance of classifiers such as SVM or fully connected neural networks. This makes CNN-based feature extraction a better choice for complex image classification tasks like bone fracture detection.

CNNs are designed to handle large image datasets efficiently. They reduce the dimensionality of images through pooling layers, which helps in managing computational complexity while preserving essential information

Classification

KNN

K-Nearest Neighbors is a supervised classification that classifies a new data point based on how its nearest neighbors are classified.
In this project we used 3,5 and 7 as the values of neighbors.



SVM

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification tasks. Its objective in classification is to find the optimal hyperplane that separates different classes of data points. In the case of bone fracture classification, SVM can be particularly effective due to its ability to handle complex decision boundaries and its robustness in high-dimensional spaces.
It provides a robust framework to differentiate between fractured and non-fractured bones, contributing to diagnostic decision-making in medical practice.

SHALLOW NEURAL NETWORKS

A shallow neural network with one hidden dense layer is trained for classification. This network includes a Flatten layer to convert the image into a vector and a Dense layer for classification.
A shallow neural network with a single hidden dense layer and a Flatten layer is a feasible and effective approach for bone fracture classification. It leverages the network's ability to automatically learn relevant features from image data, facilitating accurate classification which is crucial in medical diagnostics and treatment planning.

CNN

CNNs represent a powerful tool for bone fracture classification by leveraging their ability to automatically extract and interpret meaningful features from medical images. Their application in this domain can significantly enhance diagnostic capabilities and improve patient care through more accurate and timely identification of bone fractures.

RANDOM FOREST

Random forest is a supervised machine learning algorithm used for classification and regression.
It operates by building multiple decision trees during training and outputs the class.
Random forests are particularly advantageous for handling large and complex datasets.
They can capture complex relationships between features and labels making the model robust to noise and overfitting.

MODEL EVALUATION

When building machine learning models the goal is to make predictions based on data. To understand how well these data models perform we need to evaluate the using various metrics.

Accuracy

This metric calculates the proportion of prediction that the model got right out of all the predictions it made.
So, how often were our models right?

MODEL	ACCURACY (validation)	ACCURACY (test)
KNN k=3	0.8587 = 86%	0.8422 = 84%
KNN k=5	0.8820 = 88%	0.8311 = 83%
KNN k=7	0.8680 = 87%	0.8373 = 83%
Shallow neural network	0.8991 =90%	0.8410 = 84%
CNN	0.9767=83%	0.8708 = 87%
SVM	0.8960=89%	0.8522=85%
Random Forest (gini)	0.8696=87%	0.8248=82%
Random Forest (entropy)	0.8711=87%	0.8174=82%

Precision

Precision is a metric that measures how often a machine learning model correctly predicts a positive class.

How often the positive prediction in our models where actually correct?

MODELS	PRECISION (validation)	PRECISON (test)
KNN k=3	0.5316=53%	0.6444=64%
KNN k=5	0.6471=65%	0.6000=60%
KNN k=7	0.5902=59%	0.6528=65%
SVM	0.8993=90%	0.8446=84%
Shallow neural network	0.6703=67%	0.6016=60%
CNN	0.7534=75%	0.7882=79%
Random Forest (gini)	0.8750=87%	0.7308=73%
Random Forest (entropy)	0.9333=93%	0.6500=65%

Recall

Recall is a metric that measure how often a machine learning model correctly identifies positive instances from all the actual positive samples in the dataset.

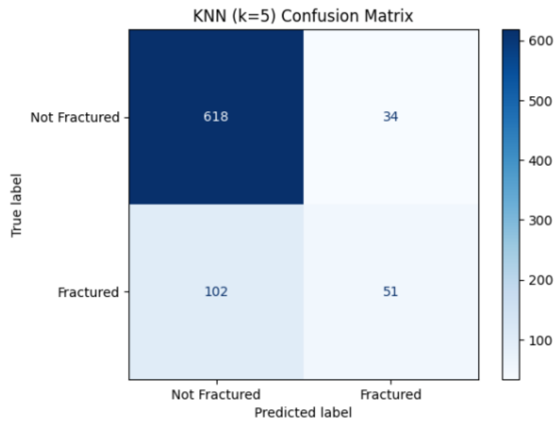
Can our models find all instances of the positive cases?

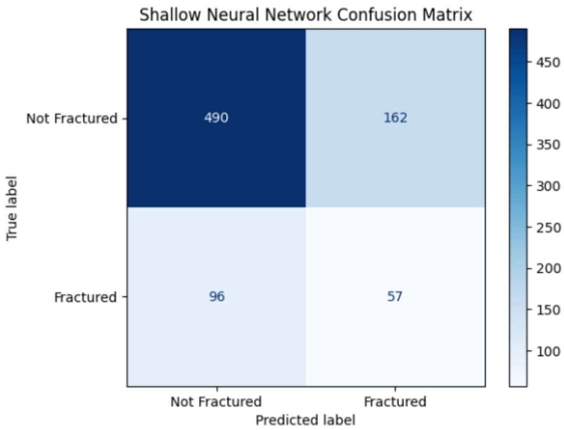
MODELS	RECALL (validation)	RECALL (test)
KNN k=3	0.4375=44%	0.3791=38%
KNN k=5	0.4583=46%	0.3333=33%
KNN k=7	0.3750=37%	0.3072=30%

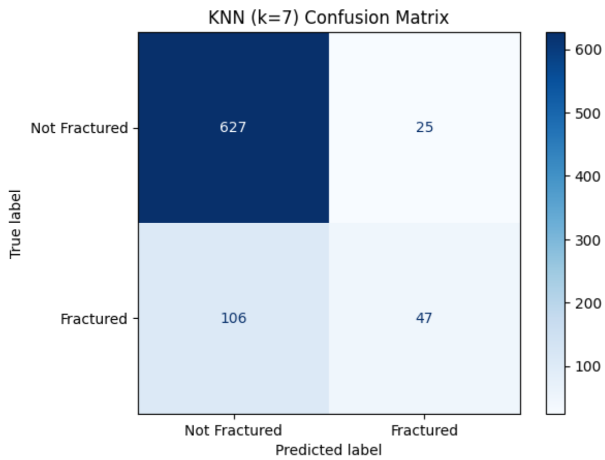
SVM	0.8960=89%	0.8522=85%
Shallow neural network	0.6354=63%	0.4837=48%
CNN	0.5729=57%	0.4379=44%
Random Forest (gini)	0.1458=15%	0.1242=12%
Random Forest (gini)	0.1458=15%	0.0850=8%

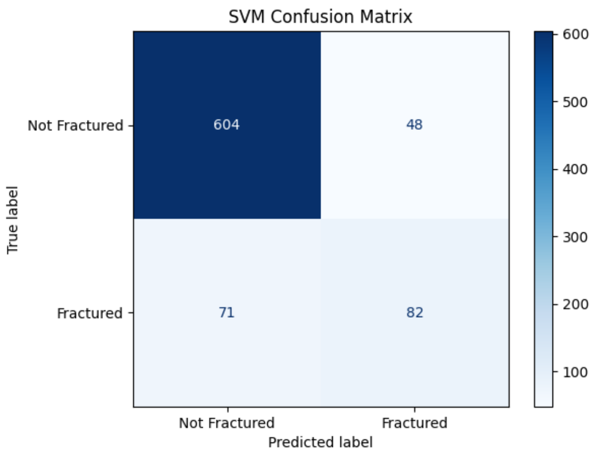
Confusion Matrix

A confusion matrix is a table that summarizes the performance of classification model by comparing its predicted labels to the true labels. It provides a detailed breakdown of the model's prediction. It displays the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) of the model's predictions.









CONCLUSION

This project evaluates the performance of several machine learning and deep learning models for the classification and localization of bone fractures, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Random Forests. Among the KNN models, K=5 achieved the highest validation accuracy at 88%, but its test accuracy dropped to 83%, indicating overfitting. Its precision and recall were relatively low, with K=3 achieving the highest test precision at 64% but only 38% recall, indicating it often failed to identify fractures correctly. The shallow neural network performed better, with a validation accuracy of 90% and test accuracy of 84%, but it struggled with precision (60%) and recall (48%) on the test set. The CNN model showed significant improvement, achieving 83% validation accuracy and 87% test accuracy, with better precision (79%) and recall (44%) compared to the shallow neural network, although its recall was still low. The SVM model provided a balanced performance with 89% validation accuracy and 85% test accuracy, coupled with strong precision (84%) and recall (85%), making it a reliable choice for fracture classification. The Random Forest models showed good validation accuracy around 87% but lower test accuracy around 82%, with varying precision and poor recall, indicating they were not as effective for this task.

Overall, while the CNN model showed the highest test accuracy at 87%, its lower recall suggested it missed many fractures. In contrast, the SVM model, with its balanced precision and recall, emerged as the best-performing model for this project. It achieved robust performance across all metrics, making it suitable for practical deployment in medical diagnostics where both detection accuracy and the ability to correctly identify fractures are crucial.

BIBLIOGRAPHY

Ahmed, K. D. (2023). *Detection of bone fracture based on machine learning techniques.* Tratto da https://www.sciencedirect.com/science/article/pii/S2665917423000594

Avinash Vishnu V, D. P. (2017). *Detection and Classification of Long Bone fractures* .