International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 247-254
https://doi.org/10.70917/ijcisim-2026-2516
Article

BINARY CLASSIFICATION OF SPINAL CORD INJURY USING ENSEMBLE MACHINE LEARNING VOTE ALGORITHMS

Jyoti. M. Waykule[1], M.P. Mattada[2], Santosh M.Nejakar[3]

[1]Sanjay Ghodawat University, Kolhapur, India., jyoti.waykule@ec.sguk.ac.in
[2]PESITM, Shivamogga, Karnataka, Belgavi, India E-Mail: mpmathad@gmail.com
[3]Jain Institute of Technology, Davanagere , Belgavi, India E-Mail: nejakarganu@gmail.com

Abstract: Spinal cord injury (SCI) is a clinically pertinent disorder which is needed to be evaluated in a timely manner. The proposed research is based on the machine learning model used for binary classification of SCI using MRI images. There are four segmentation (Fuzzy C-Means, Region Growing, K-Means and Expectation Maximization) that were evaluated and K-Means was the best performer due to its effectiveness in identifying the spinal cord structures. GLCM texture attributes were also extracted on the segmented areas where 14 features had been extracted and narrowed down to 8 features to be applied in the categorization. Three classical machine learning models were built and tested on the identical set of features. The ML models utilized were: 1) K-Nearest Neighbour (KNN), 2) Probabilistic Neural Network (PNN) and 3) Naive Bayes (NB). It was found that KNN was the most balanced to the overall performance. The findings indicate that conventional classifiers, with the application of the features based on textures, could be useful in assisting the detection of automated SCI. The study provides a rational and understandable basis that future solutions grounded on hybrids or ensembles can be created with the aim to improve the accuracy of diagnosis.
Keywords: NA

1. INTRODUCTION

1.1 Background and Motivation

Spinal cord injury (SCI) is a severe clinical disease, which may cause partial or complete sensory and motor impairment. Proper and early detection of symptomatic SCI is important since prompt treatment could minimize the neurological impairment observed in the long-term and enhance recovery of the individual. In medicine, magnetic resonance imaging (MRI) is the procedure of choice when imaging the spinal cord due to the ability to image soft tissues with clarity and detect structural abnormalities including edema, haemorrhage, and compression of the cord.

Even though MRI produces quality information, it requires strong experience of the radiologist to interpret it. SCI in simpler spinal scans may go undetected particularly when the image contrast is poor or the injury limit is not distinct. Different experts also have different interpretations of the human interpretation and this might result in lack of consistency in diagnostic decisions. Due to this, computer-aided detection (CAD) systems, which can assist radiologists through providing objective and repeatable evaluation of spinal cord image, have become of growing interest. CAD systems are developed based on the image processing and machine learning to identify patterns related to injury, and may serve as a second reader to clinical decision-making. Fig 1.1 provides an illustration of spinal cord injuries.

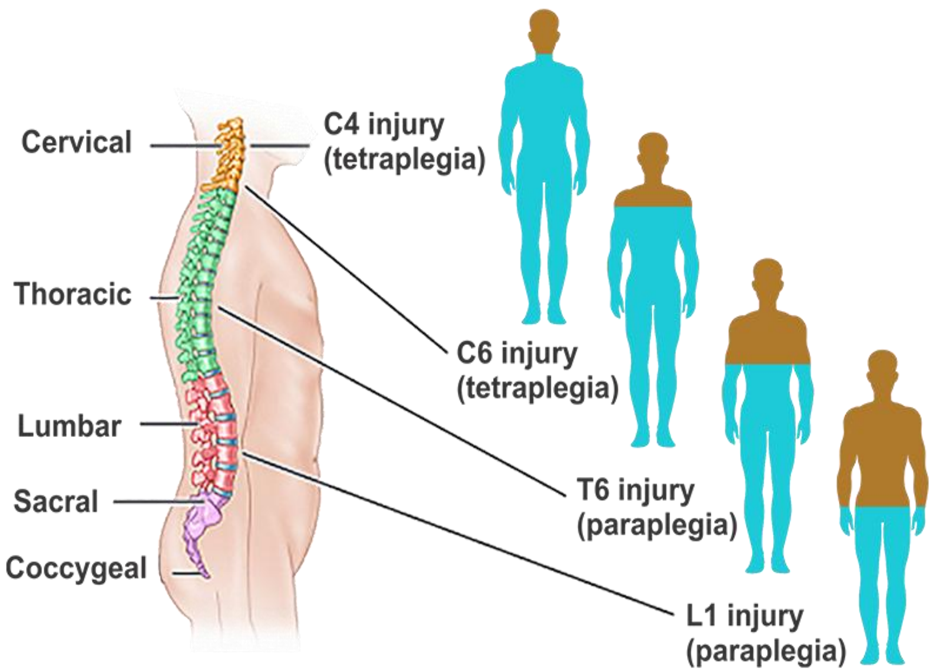

Fig 1.1. Illustration of spinal cord injury

Source: https://www.neurogen.in/spinal-cord-injury

1.2 Imaging Challenges in SCI Classification

The images of the spine cord MRI pose a number of challenges to the automated analysis. The spinal cord is a long and thin structure that is surrounded by cerebrospinal fluid, vertebrae and soft tissues which tend to present with similar intensities. This complicates the process of boundary detection. Most MRI slices are also noisy and have poor contrast, and limited anatomic variation with different individuals. Such variations diminish the performance of the classification models in case the images are not properly segmented.

The significance of segmentation in SCI detection is that it isolates the spinal cord area and minimises the effect of the background structures. A good segmentation technique will make sure that the henceforth characterization procedure will obtain the characteristics of the spinal cord. The lack of adequate segmentation can result in the classifier acquiring patterns that are not relevant to the injury so that the classification performance is poor. This is the reason why it is necessary to evaluate various segmentation techniques before classifying it.

1.3 Problem Definition

The article is concerned with binary classification of spinal cord MRI samples that are either injured or non-injured. It is intended to analyzing the performance of three machine learning classifiers through K-Nearest Neighbour (KNN), Probabilistic Neural Network (PNN) and Naive Bayes (NB) trained using consistently segmented and reduced feature MRI data. The features of all classifiers are based on the same set of features based on the texture, and the same segmentation result is used to make the comparison as fair as possible.

The objective is largely to find the classifier that gives the most accurate prediction of SCI. The comparison assists in learning the behaviour of different lightweight and interpretable models when used with medical image data.

1.4 Contributions of This Work

This paper gives a full and coherent pipeline of SCI classification. The significant contributions are as follows:

- Four segmentation methods are used in spinal cord MRI images (Fuzzy C-Means (FCM), Region Growing, K-Means, Expectation Maximization (EM)).
- A comparative evaluation of these techniques is conducted when 3 machine learning classifiers such as KNN, PNN and Naive Bayes are trained using the same feature set.
- A complete comparison is presented using accuracy, sensitivity, specificity, precision, recall, and F-score.

This experimental assessment illustrates the significance of the quality of segmentation, feature selection and behaviour of the algorithm where medical image data are involved.

2. LITERATURE REVIEW

Machine learning has attracted a lot of interest towards the field of spinal cord injury (SCI), and several studies have tried to forecast functional outcomes, detect risk patterns, or aid in clinical decision-making. The clinical variables prediction of rehabilitation progress can be regarded as one of the first lines in this direction. An ensemble machine learning model was built by a study in a spinal center, Tokyo, to predict the functional independence scores at discharge using admission-time features and the meta-model was better on the accuracy and error measures than individual base learners [1]. Though this indicates the high predictive usefulness of clinical characteristics, it does not analyze MRI values or compare basic classifiers in an image-based setting.

A lot of similar works have been dedicated to predict SCI among patients with cervical spondylosis with a multicenter dataset. In one such study, LASSO regression was used to select eleven core predictors, and the results of a ten-machine learning models comparison revealed that the random forest model was the most accurate and possessed the highest AUC in all the evaluation sets [2]. This study supported the usefulness of ML in clinical diagnosis, but, similar to the last one, it is based mostly on the structured tabular data as opposed to the image segmentation or texture-based classification.

Motor recovery prognostication has also been studied using advanced modelling. Nested ensemble algorithm has been developed in [3] to forecast six-month motor score results in traumatic cervical SCI patients using only the very early ASIA motor score. The two-stage architecture obtained an accuracy of approximately 80% and showed balanced sensitivity and specificity which implies a great possibility of the early clinical decision support. Nonetheless, it does not touch upon the imaging field and targets instead numerical clinical scales.

The use of tree-based algorithms has also been thoroughly studied in functional outcome prediction. A retrospective study of 589 rehabilitation patients compared various machine learning models and concluded that random forest and XGBoost made the best predictions of Functional Independence Measure (FIM) scores at discharge, which is much better than generalized linear models [4]. Once again, the findings are encouraging in regard to estimating the outcome, but the research does not examine MRI processing, segmentation, and texture-based characteristics, which form the basis of image-level SCI detection.

Disease-specific prediction of SCI has also been done using machine learning. A multicentric study on spinal tuberculosis patients had ten significant clinical predictors and established that the random forest was the best predictive model in forecasting the onset of SCI, with the external AUC of 0.816 [5]. This work is not concerned with imaging of the spinal cord itself, although the parameters of the clinical aspects and systemic disease are relevant to the clinical trial.

In addition to the individual modeling work, a systematic review of the increasing use of ML in traumatic spinal cord injury has been summarized in [6]. A PRISMA-based review concluded that ML has been utilized in diagnosis, prognostication, rehabilitation planning, and chronic complications monitoring with such algorithms as SVM, ANN, CNN, and decision trees being frequently seen in applications. The review also identified the natural variability of injury progression as a good rationale behind ML approaches but indicated that there were limited consistent studies that had used imaging to diagnose the injury. This is in line with the rationale behind the current work that focuses particularly on features based on MRI.

Subsequent research has tried to do extensive clinical prognostication clinically on intensive care data. An analysis of 1485 ICU cases by machine learning produced 98 classifiers that predicted the destination of discharge and it was found out that an ensemble model was more discriminating, prediction consistent and beneficial in decision curves than the single classifiers [7]. There are similar attempts that utilized large ICU databases to forecast long ICU stay, and long hospital stay, where ensemble models once again indicated the most significant AUC scores in both internal cross-validation and independent analysis [8]. These works illustrate the high role of ML in management of resources at hospital level but fail to accommodate the imaging-based diagnosis.

A systematic review [9] summarized evidence of 21 studies and found that machine learning and deep learning technologies can predict outcomes, survival, and neurological evaluations in SCI, nonetheless, more refined diagnostic tools should also be developed, in particular, the ones based on imaging modalities. This fact justifies the gap to be discussed in the current research, i.e., to segment, extract features, and compare classifiers using MRI instead of clinical variables.

Lastly, secondary complications of SCI are the targeted condition used in machine learning. An investigation was done on the postoperative pulmonary infection where 6 ML models were utilized, and the random forest model showed the best predictive performance which was measured by the AUC, accuracy, and sensitivity values that are reliable predictors [10]. Despite being helpful in managing the postoperative period, this work is silent on the imaging of the spinal cord and classification of injuries.

Identified Research Gaps

1. The majority of current SCI machine learning research is based on clinical variables, demographic information, or rehabilitation data, and there is a very less research where MRI images are taken as the main input to identify the presence of an injury.
2. Not a lot of research combine a variety of segmentation methods prior to classification, despite the fact that the quality of segmentation has a direct impact on the precision of image-based diagnosis.
3. Simple and interpretable classifier, including KNN, PNN, and Naive Bayes, have not been comparatively analyzed.
4. No single and reproducible pipeline exists which includes segmentation, feature extraction, feature selection and comparison of classifiers on the same standardized dataset.

This paper constructs a full MRI work flow through four types of segmentation, GLCM texture features extraction, and KNN, PNN, and Naive Bayes evaluation under the same condition. In this way it gives a transparent, easy to use, and repeatable pipeline to binary classify spinal cord injury based on image-derived features instead of clinical information.

3. METHODOLOGY

This suggested binary classification workflow to detect spinal cord injury (SCI) on MRI images has eight stages, as shown in Fig. 3.1. The diagram is divided into two horizontal rows, the upper one is the imaging and feature-engineering workflow, and the lower one is the machine learning and evaluation workflow. All stages make a radical change in the information to guarantee credible and interpretative classification [11].

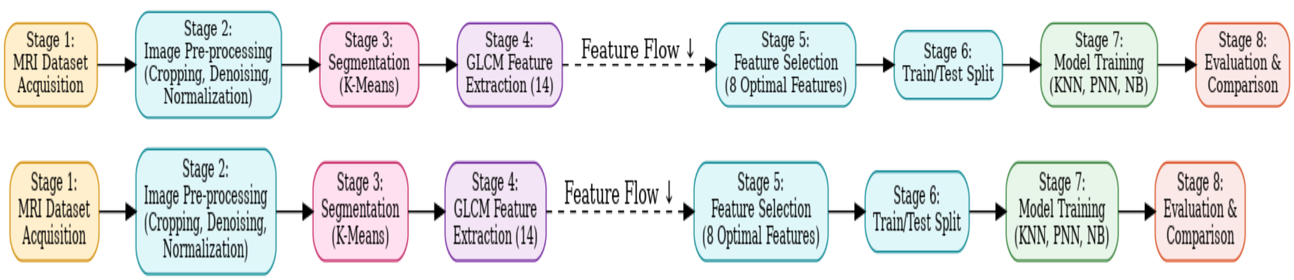

Fig 3.1: Proposed Flow

Stage 1: MRI Dataset Acquisition: MRI images of the spinal cord were obtained from an online repository, and only those were taken into account that had the cord clearly shown in the slice. The dataset for the experiment was downloaded from http://vislab.gtu.edu.tr.The T1- and T2-weighted mid-sagittal lumbar MR images from 80 patients are included in the collection. Radiological description was used to identify each sample as injured or non-injured [12]. The objective is to construct clean data with uniform quality of imaging and true ground truth.

Stage 2: Image Pre-processing: In order to normalize the images, four processing were used: spinal-area cropping, denoising [13], which is used to remove random noise, image intensity normalization, which is used to reduce scanner-specific variation, and contrast enhancement (CLAHE) to emphasize the anatomical boundaries. These measures normalize the input of clustering-based segmentation.

Stage 3: Segmentation (K-Means): The images obtained after stage 2 were pre-processed followed by the use of K-Means clustering to isolate the spinal cord area. K-Means was selected as they are simple and have a consistent boundary recovery in this data. The segmentation result is used to specify the region of interest onto which the texture measures are to be obtained.

Stage 4: GLCM Feature Extraction: GLCM texture features were calculated at various orientations in a given segmented spinal cord region yielding fourteen features. These properties characterize spatial grey-level pattern which are associated with injury-related structural changes. The feature vectors that have been extracted are used as brief numerical classifier inputs.

Stage 5: Feature Selection & Scaling: Correlated features or features with low-discrimination were cut and 8 best features were retained. All the features were then scaled (normalized) to avoid scale imbalance, which is significant to distance-based (KNN) and probability-based (PNN, NB) models.

Stage 6: Train/Test Split: Stratified sampling was used to split the dataset into a training and a testing subsample to maintain the ratio of classes. This makes sure that the assessment indicates the actual performance on unseen images.

Stage 7: Model Training (KNN, PNN, NB): Three traditional classifiers: KNN, PNN and Naive Bayes trained on the chosen features.

- KNN uses neighbourhood distance patterns [14].
- PNN estimates class-conditional probability densities [15].
- NB computes posterior probabilities under feature independence [16].

It is possible to compare all the models during training on the same features.

Stage 8: Evaluation & Comparison: Accuracy, specificity, precision, sensitivity, and F1-score were used to evaluate classifier predictions. Errors were inspected though confusion matrices. KNN also performed best in general under the same conditions and input.

4.RESULTS

The three classifiers KNN, PNN, and Naive Bayes (NB) were tested using the eight texture features selected as GLCM. Diagnostic capability was evaluated by use of accuracy, Sensitivity, Specificity, Precision, Recall, F1-score, and False Positive Rate (FPR). All the comparative results are presented in Table 4.1.

Metric	KNN	PNN	NB
Accuracy	79.17%	80.83%	73.75%
Sensitivity	69.57%	69.57%	76.81%
Specificity	92.16%	96.08%	69.61%
TPV (Precision)	92.31%	96.00%	77.37%
FPR	7.84%	3.92%	30.39%
Recall	69.57%	69.57%	76.81%
F1-Score	79.34%	80.67%	77.09%

Table 4.1. Performance of KNN, PNN, and Naïve Bayes

The bar plots below compare the metrics as shown.

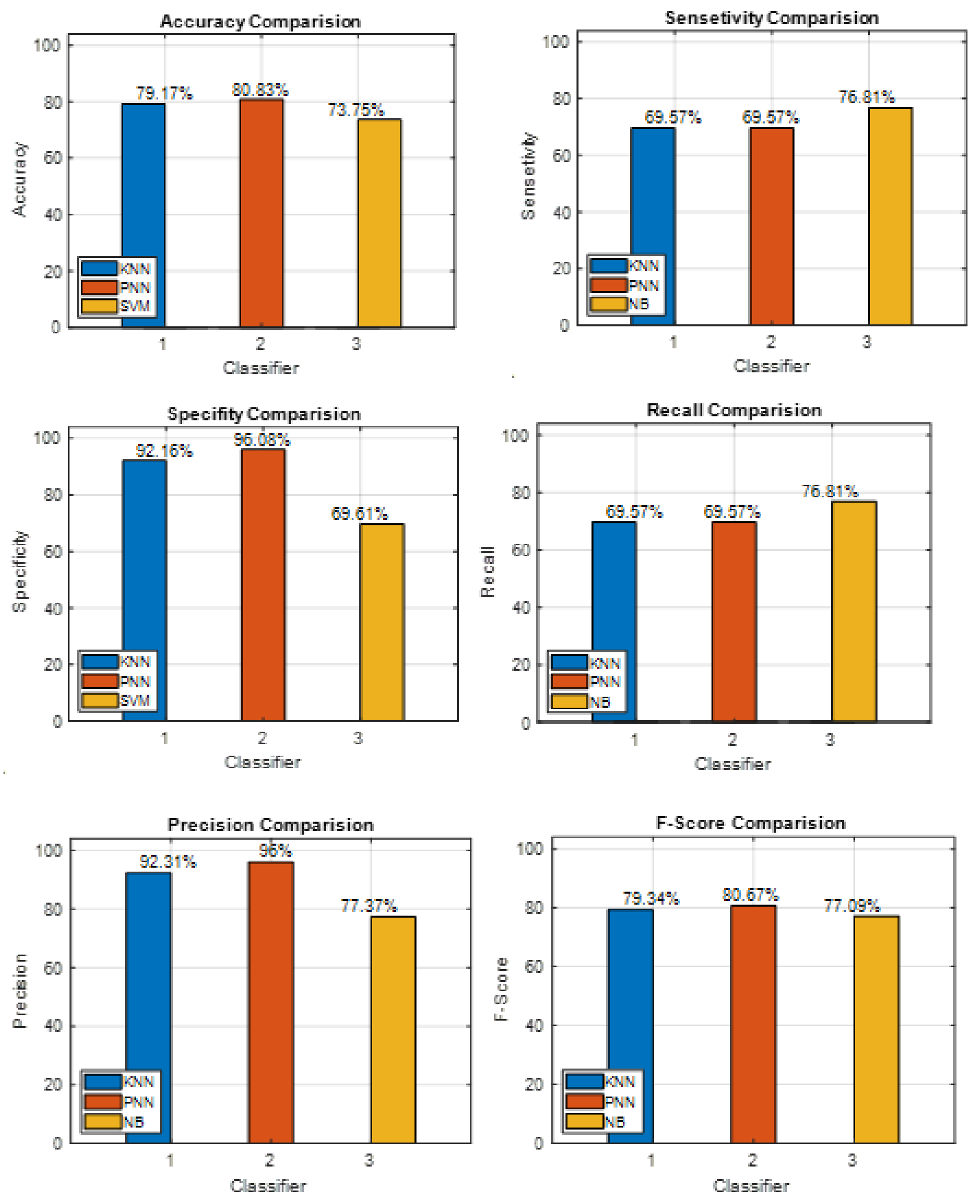

Fig 4.1: Bar plot comparison for the evaluation metrics

The comparative analysis of KNN, PNN and Naive Bayes indicates unique behavioural patterns indicating interaction manner of each of the models with the GLCM texture feature space. PNN has the best accuracy (80.83%), precision (96.00%), and specificity (96.08%) which implies that it has good capability of approximating the distribution of non-injured tissue with the least number of false positives. PNN is good at learning constant texture patterns, which are assisted by its smooth probability surfaces created by its radial basis functions. KNN was also competitive with balanced accuracy (79.17) and high precision (92.31) which can be attributed to its neighbourhood-based decision-making mechanism. These findings indicate that the texture characteristics in the feature space are organized in local clusters, which KNN can effectively classify most non-injured cases, and have tolerable sensitivity.

Naive bayes showed the best level of sensitivity (76.81%), that is, it was more effective than the other classifiers at identifying the injured cases. Though, it was obtained at the expense of low specificity (69.61) and false positive rate is much higher (30.39). Such behaviour is in line with the assumption of independence of NB that is violated in GLCM features that are inclined towards being correlated. Consequently, NB inflates the risk of injury, and this increases false alarms. Although less accurate than the other, the sensitivity of NB also shows that it can be useful in the initial detection of injuries, whereas other models such as PNN and KNN offer more reliable and conservative confirmation.

Comprehensively, the KNN balanced metric profile and the PNN high specificity justify that classical machine learning models that are based on texture can be effective in cases where they are consistent with the statistical aspects of the features.

5. DISCUSSION

5.1 Summary and Interpretation of Findings

The paper has contrasted the performance of KNN, PNN and Naive Bayes classifier in binary differentiation of spinal cord injury using GLCM texture features that are extracted on K-Means segmented MRI images. The results of the evaluation indicate different behavioural patterns of the three algorithms, the way the various models act with the underlying feature space.

PNN had the greatest overall accuracy and precision. This can be anticipated since PNN employs radial basis functions in approximating class-conditional densities thereby enabling it to depict smooth decision boundaries. The texture of the spinal cord obtained with the help of GLCM matrices are inclined to continuous distributions, and PNN is able to detect these relationships. It's very low false positive rate and its high specificity suggest that PNN is especially useful in indicating healthy tissue without excessively predicting the injury. Nevertheless, its sensitivity was moderate which implies that the model can be conservative in terms of flagging cases of injuries when there is a high overlap of features.

KNN had balanced and steady results in every metric. The findings indicate that GLCM features generate some local clustered patterns in feature space, which can be used by KNN in terms of distance-based neighbourhood decisions. The fact that it performs well near PNN suggests that the dataset has a comparatively smooth feature structure, which can be used by even a simple classifier. The moderate sensitivity signifies that there are samples of the injured that lie close to the boundary areas yet KNN still yields a high precision and specificity thus it is reliable when using it in clinical settings where the false positive should be at a minimum.

Naive Bayes demonstrated the greatest sensitivity, which means that it is more aggressive when it comes to the detection of injured tissue. This however came at a high price of reduced specificity and precision. This behaviour is consistent with the assumption of feature independence by NB which is not true in the case of correlated texture features of GLCM that are correlated in nature. Consequently, NB gives too many chances of an injury happening resulting in a very high false positive. This is the reason why it is less accurate overall and has lower reliability as a diagnostic model.

Altogether, KNN balanced behaviour demonstrates the appropriateness of texture-based ML methods to MRI classification. The findings also indicate the relevance of choosing models that correspond to feature behaviour particularly in medical imaging procedures.

5.2 Limitations of the Work

Several limitations should be acknowledged when interpreting the results of this study:

1. Restricted Dataset Size: The dataset comprises a limited number of MRI slices, which may not capture the complete anatomical variability seen in clinical environments [17].
2. Limited Evaluation Metrics: The study focused on conventional metrics but did not explore advanced evaluations [18-20].

These limitations indicate areas where methodological expansion could strengthen the reliability and generalizability of the system.

5.3 *Future Scope*

The study can be developed further in future by considering more sophisticated modelling strategies and more specifically hybrid ones that would integrate strengths of several classifiers. These models could make use of the balanced sensitivity-specificity trade off and resistance to non-linear local trends of KNN. The complementary behaviours could be combined in a hybrid ensemble or stacked model, where each of the classifiers is allowed to make selective contributions to the ultimate predictions.

6. CONCLUSION

This work outlined an entire workflow of binary classification of spinal cord injury with the help of MRI segmentation and machine learning. The original segmentation techniques investigated are four and K-Means was chosen because it detects anatomical boundaries with consistency. Out of the segmented regions of the spinal cord, fourteen GLCM texture features were obtained and eight best features were used in the classification process. There were three classical machine learning models, namely KNN, PNN, and Naïve Bayes that were tested under the same conditions.

The findings indicate that PNN attained the best accuracy, precision and specificity whereas Naïve Bayes had the best sensitivity with a high false positive rate. KNN has provided the best-balanced result, with high specificity and precision, and F1-scores were also stable. This balance renders KNN a solid and understandable model to identify any spinal cord injury using the texture features.

Comprehensively, it can be affirmed that machine learning models with lightweight coupled with strong segmentation and texture extraction are capable of assisting automated SCI identification. The gained insights are also used to point out the possibility of hybrid models, which can be used to combine the strengths of single classifiers to improve diagnostic reliability in the future work.

References

1. Kato, Chihiro, et al. "Functional outcome prediction after spinal cord injury using ensemble machine learning." Archives of physical medicine and rehabilitation 105.1 (2024): 95-100.
2. Zhou, Z., Feng, S., Zhou, X., Yu, J., Zhu, J., Pang, L., ... & Liu, J. (2025). Using multiple machine learning algorithms to predict spinal cord injury in patients with cervical spondylosis: a multicenter study. Scientific Reports, 15(1), 37418.
3. Wang, Yijin, et al. "Application of a novel nested ensemble algorithm in predicting motor function recovery in patients with traumatic cervical spinal cord injury." Scientific Reports 14.1 (2024): 17403.
4. Rasoolinejad, Mohammad, et al. "Machine learning predicts improvement of functional outcomes in spinal cord injury patients after inpatient rehabilitation." Frontiers in Rehabilitation Sciences 6 (2025): 1594753.
5. Feng, Sitan, et al. "Prediction model for spinal cord injury in spinal tuberculosis patients using multiple machine learning algorithms: a multicentric study." Scientific Reports 14.1 (2024): 7691.
6. Dietz, Nicholas, et al. "Machine learning in clinical diagnosis, prognostication, and management of acute traumatic spinal cord injury (SCI): A systematic review." Journal of Clinical Orthopaedics and Trauma 35 (2022): 102046.
7. Fan, Guoxin, et al. "Early prognostication of critical patients with spinal cord injury: A machine learning study with 1485 cases." Spine 49.11 (2024): 754-762.
8. Fan, Guoxin, et al. "Machine learning-based prediction of prolonged intensive care unit stay for critical patients with spinal cord injury." Spine 47.9 (2022): E390-E398.
9. Habibi, Mohammad Amin, et al. "Predicting the outcome and survival of patients with spinal cord injury using machine learning algorithms: a systematic review." World Neurosurgery 188 (2024): 150-160.
10. Li, Meng-Pan, et al. "Machine learning for the prediction of postoperative nosocomial pulmonary infection in patients with spinal cord injury." European Spine Journal 32.11 (2023): 3825-3835.
11. Waykule, Mrs Jyoti M., and V. R. Udupi. "Spine Image Segmentation using Fuzzy C-Mean."
12. Waykule, Jyoti M., and V. R. Udupi. "Spine Diseases Detection Using SVM." Techno-Societal 2020: Proceedings of the 3rd International Conference on Advanced Technologies for Societal Applications—Volume 1. Cham: Springer International Publishing, 2021.
13. Waykule, Jyoti M., et al. "A Comparative Study of Clustering Algorithms for MRI Spine Image Segmentation: Fuzzy C-Means, Region Growing, K-Means and Expectation-Maximization." Journal of Information Systems Engineering and Management, vol. 10, no. 48s, 2025, pp. 452–460.
14. Maki, Satoshi, et al. "Machine learning and deep learning in spinal injury: a narrative review of algorithms in diagnosis and prognosis." Journal of Clinical Medicine 13.3 (2024): 705.
15. Fallah, Nader, et al. "Development of a machine learning algorithm for predicting in-hospital and 1-year mortality after traumatic spinal cord injury." The spine journal 22.2 (2022): 329-336.
16. Kapoor, Dhruv, and Clark Xu. "Spinal cord injury AIS predictions using machine learning." eneuro 10.1 (2023).
17. Kapoor, Dhruv, and Clark Xu. "Spinal cord injury AIS predictions using machine learning." eneuro 10.1 (2023).
18. DelSole, Edward M., Wyatt L. Keck, and Aalpen A. Patel. "The state of machine learning in spine surgery: a systematic review." Clinical Spine Surgery 35.2 (2022): 80-89.
19. Shimizu, Tomoaki, et al. "Efficacy of a machine learning-based approach in predicting neurological prognosis of cervical spinal cord injury patients following urgent surgery within 24 h after injury." Journal of Clinical Neuroscience 107 (2023): 150-156.
20. Karabacak, Mert, and Konstantinos Margetis. "Precision medicine for traumatic cervical spinal cord injuries: accessible and interpretable machine learning models to predict individualized in-hospital outcomes." The Spine Journal 23.12 (2023): 1750-1763.