

Design of a novel deep network model for spinal cord injury prediction

P. R. S. S. Venkatapathi Raju^{1,2}, Valayapathy Asanambigai¹, Suresh Babu Mudunuri²

¹Department of Computer Science Engineering, Annamalai University, Annamalaiagar, India

²Department of Information Technology, Sagi Rama Krishnam Raju Engineering College, Bhimavaram, India

Article Info

Article history:

Received Apr 6, 2023

Revised Sep 11, 2023

Accepted Dec 2, 2023

Keywords:

Accuracy

Convolutional network

Deep learning

Prediction

Spinal cord injury

ABSTRACT

Degenerative cervical myelopathy must be diagnosed with magnetic resonance imaging (MRI) which predicts spinal cord injury (SCI). The growing volume of medical imaging data can be managed by deep learning models, which provide a preliminary interpretation of images taken in basic care settings. Our main goal was to create a deep-learning approach that could identify SCI using MRI data. This work concentrates on modeling a novel 2D-convolutional neural networks (2D-CNN) approach for predicting SCI. For holdouts, training, and validation, various datasets of patients were created. Two experts assigned labels to the images. The holdout dataset was used to evaluate the performance of our deep convolutional neural network (DCNN) over the image data from the available dataset. The dataset is acquired from the online resource for training and validation purpose. With the available dataset, the anticipated model attains 94% AUC, 0.1 p-value, and 92.2% accuracy. The anticipated model might make cervical spine MRI scan interpretation more accurate and reliable.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

P. R. S. S. VenkatapathiRaju

Department of Computer Science Engineering, Annamalai University

Km. 10 Annamalaiagar, India

Email: prssvraju2@gmail.com

1. INTRODUCTION

Spinal cord injury (SCI) prediction is extremely complex. Hence strong statistical methods and efficient clinical predictive models (CPMs) must be developed [1]. CPMs establish a connection between a desired result, statistical rules, a predictor variable, or variables. Clinicians commonly use CPMs to determine the best course of treatment, manage patient expectation and anticipate an illness's course [2]. Pathophysiological alterations on both traumatic and non-traumatic SCI cause glial scarring and the creation of cystic cavities, which prevent neuronal regeneration and repair [3]. Clinically, this seems a significant, maybe permanent functional, decline that affects people, families, and society as a whole [4].

Machine learning (ML) is an acronym for a group of computer algorithms that may now find connections between datasets used more and more in creating this CPMs [5]. ML has many advantages over conventional predictive models that incorporate logistic regression in some way [6]. First, because it frequently uses the dataset automatically, ML rarely needs prior knowledge of significant predictors. Second, ML typically has fewer limitations than most predictors that can be used in a given dataset with logistic regression [7]. Therefore, ML is advantageous for huge datasets (like those in cancer and pharmacogenomics), where many predictors are present but not always clear links between them [8]. Due to these advantages, it is usually discovered that ML is more precise and effective than logistic regression approaches when applied to the same dataset. This comparative advantage has increased the usage of ML in

SCI predictive modeling to the idea that several research teams have just started to develop ML-based CPMs for SCI [9]. Thirdly, ML can uncover intricate, nonlinear correlations in datasets that logistic regression is less suited to creating and analyzing [10].

In supervised ML, the user labels the data's inputs and outputs before forming the model [11]. In other words, by labeling the data in advance, the user effectively "supervises" the algorithm, using ML to build a model that links inputs and outputs [12]. Unsupervised ML does not label the input or output of the data. Instead, the algorithm searches for characteristics in the data that let it combine various data pieces. The algorithm in reinforcement learning looks at its surroundings to find the action that would maximize reward. As an example, a robot that has been taught to play tennis is taught to choose moves in a certain situation that maximize decreasing the penalty (total points) while increasing reward (total points) (i.e., tennis court) [13]. In SCI, supervised or unsupervised learning are more suitable for using epidemiological datasets than reinforcement learning. Supervised learning models, which incorporate regression and classification techniques, make up the vast bulk of ML models created in SCI [14].

Training dataset is used to build and optimize an ML model, while a different collection of testing data is utilized to build supervised ML techniques. An illustration of the train test split is shown. An effective ML model must maintain a training-testing split because it guards against overfitting and offers an exhaustive preliminary assessment of the model's external validity. It is crucial to understand that ML is not a single, all-encompassing concept; rather, ML models can be created using various algorithms. Support vector machines (SVMs), naive Bayes (NB), decision tree (DT), and k-nearest neighbors (KNN) are examples of these supervised methods. Additionally, ensemble approaches like bagging and boosting can be used to modify ML models further. For instance, various subsets of the training data can be used to train distinct classification trees [15]. The final step is to create a random forest model by setting the combined majority result of the individual classification trees as the result of bagging the classification trees. Stacking may be used by similarly mixing the results of completely separate ML models. ML is the preferred tool for creating high-performance predictive models due to its variety and computing capacity. Furthermore, the nonlinear correlations that would be extremely challenging to incorporate in a straightforward logistic regression can be found because of the structure of these models. The major research contributions are:

- a. The dataset is taken from the online available resources, i.e. SCI-based magnetic resonance imaging (MRI) dataset and pre-processing is performed to eliminate the noise of the model using gray matter (GM) analysis;
- b. Then feature learning is done with the proposed convolutional neural networks (CNN) model where the proposed network works with two phases, i.e. 2D-CNN slice level and 3D-CNN subject level. Here, the slicing network performs learning rate and batch normalization.
- c. The simulation is done in MATLAB 2020a environment where various metrics like training and validation accuracy, hold out, sensitivity and precision are evaluated and compared with other approaches. The model establishes better trade-off compared to the given approaches.

This review addresses recent studies that developed predictive models based on ML to forecast SCI outcomes. The following is the order: the methods utilized to create the body of research examined in this article are presented in section 2, methodology is presented in section 3. ML provides predictions for traumatic and non-traumatic SCI, as described in sections 4. Section 5 gives the work summary.

2. RELATED WORKS

Both individuals and the healthcare system are significantly impacted by traumatic spinal cord injury (tSCI). In western countries, it influences 20 and 50 individuals/million annually. The fact that affected individuals usually experience the aftereffects of chronic neurological dysfunction across different functional domains makes tSCI particularly problematic [16]. These consequences severely impact the healthcare system and cause permanent impairment. Many concerns need to be answered by ML due to the catastrophic effects of tSCI and the numerous unresolved problems surrounding its management and prognosis [17]. Due to ML's relative novelty and the propensity to focus on a more well-known regression modality instead, there is currently a relatively little body of work on its use to predict outcomes in tSCI [18].

Despite the lack of information, some important studies using numerous SCI fields have published work on using ML to improve tSCI. To identify SCI in individual axial spinal cord slices derived from diffusion tensor imaging (DTI), [19] used ML approaches DTI. The authors employed fractional anisotropy, a DTI metric that expresses the degree of tissue orientation and water diffusion in an image. It serves as a proxy for the tissue's structural integrity. Before beginning their work, the authors devised strategies to separate the spinal cord image [20]. The authors discovered a significant association between motor tests performed before and after admission and the degree of the contusion damage based on this segmentation (i.e., how much the spinal cord is that the CNN evaluated to be injured) [21]. New research using ML to

predict tSCI has achieved significant advancements in predicting functional results after SCI imaging prediction [22]. Using information from the risk hansen spinal cord injury registry (RHSCIR) dataset related to tSCI, DeVries [23] recently constructed an unsupervised ML system to predict the ability to walk independently after tSCI at discharge or after a year [23]. The performance of the unsupervised ML method and the previously developed logistic regression models, which both predicted identical outcomes, were then compared [24]. The authors discovered no appreciable differences between the two. Although this comparability could be more encouraging at this point for comparisons was established between ML and well-known statistical models [25].

ML has been used to forecast the use of opioids [26]. A damaged intervertebral disc is removed, and two nearby vertebral bodies are fused during the surgical operation known as anterior cervical discectomy and fusion (ACDF) which is used to treat dilated cardiomyopathy (DCM). Researchers found anti-depressant use, nicotine use, medicaid insurance status, and preoperative opioid prescription contributed to long-term opioid usage following ACDF. Litjens *et al.* [27] improved the applicability of their machine-learning techniques in clinical settings by developing a website to execute the model developed during the study by creating ML algorithms that predict prolonged by [28] to the lumbar domain by examining lumbar disc herniation. Instrumentation, preoperative opioid usage and concurrent depression are strongest predictors of prolonged opioid. Logistic regression model with elastic-net penalization (area under the receiver operating characteristic (ROC) curve (AUC) = 0.81 with appropriate calibration) demonstrated the best performance [29]. Recovery ratio computes the postoperative ratio for ideal enhancement. It is used to predict outcomes for imaging interpretation (pre-operative baseline recovery to maximal modified Japanese Orthopedic Association (mJOA) score). During the 1-year follow-up, patients are partitioned into two groups based on classification algorithms, whether the recovery ratio is "good" or "poor." Authors used multivariate logistic regression and supervised machine learning techniques (SVM, KNN) in their investigation. These methods made predictions regarding the recovery rate using DTI characteristics such as mean diffusivity, axial diffusivity, and fractional anisotropy (where the rate of molecule dispersion in various directions is known as diffusivity) [30]. The SVM model outperformed the logistic regression model in performance in terms of prediction accuracy; however, the deep learning (DL) model did poorly (59.2%) in terms of recovery rate prediction accuracy.

3. METHODOLOGY

This section discusses two types of convolutional network models, i.e., 2D-CNN and 3D-CNN model SCI prediction. The proposed methodology undergoes pre-processing to eradicate the artifacts and classification with the proposed idea as in Figure 1. The evaluation is done with MATLAB 2020, and various metrics like accuracy, precision, F1-score, and recall are compared and with other approaches. The analysis proves that the model works well and establishes a superior decision support system to assist the experts.

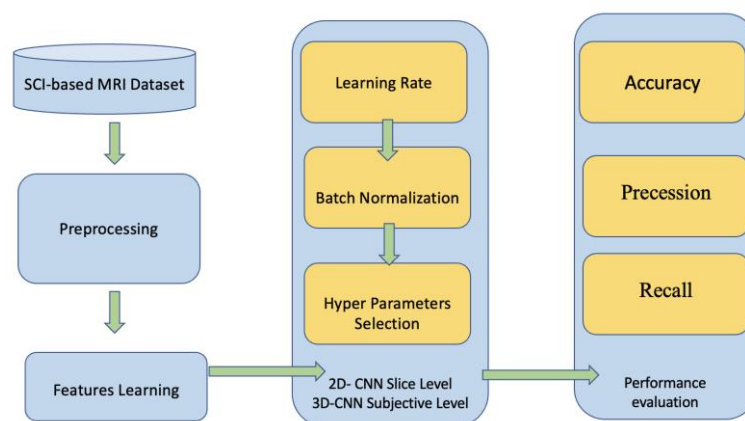


Figure 1. Proposed model's flow

3.1. Pre-processing

Here, two pre-processing methods are employed over MRI images: grey mapping was produced using pre-preprocessing. Using the N4 strategy, scans were non-uniformity corrected before being translated to spinal space by registering masks using similarity transformation. This minimum pre-processing enabled the creation of MRI images. A rigid transformation with isotropic scaling is known as a similarity

transformation. Elastix registration software was used to complete registrations. The SCI mask was altered to contain images with a zero mean and unit variance to accommodate for variations in signal strength. GM density was encoded using modulated maps. Pair-wise registration was used to establish the group template space and compute these maps without bias toward individual MRI images.

An affine, non-rigid and similarity-based transformation model was sequentially used for the pair-wise registrations. To determine the template space, we choose a portion of the images. This template set consisted of SCI patient images chosen randomly while maintaining the number of diagnostic groupings. Using the same registration process, the additional images from the data sets from the template area had registrations for SCI. The template space creation process was modified for the current work to include non-uniformity correction, skull-stripping, and template space-matched SCI space. This work adopts similarity registration with SCI masks to compute the coordination space to every image which is concatenated with average pair-wise transformation. Probabilistic GM maps are produced using statistical parametric mapping (SPM8's) unified tissue segmentation method. To create the final feature maps, which included compression and expansion, the probabilistic GM maps are transferred and multiplied using Jacobian model over deformation field. Modified GM maps are split by intracranial volume to account for image size.

3.2. Classifier

This research aims to develop a technique for detecting SCI by examining MRI spinal sequences. The spinal-based indications used in the samples are considered a reference that can be used to categorize the seriousness of SCI illness. Its range is 0 to 2, with 0.5 signifying an extremely mild condition, 1 denoting a light condition, and 2 denoting a severe condition (severe). A classification model using two, three, or four classes is created depending on the degree of precision. We combined the categories because they had fewer data than the other two classes, allowing us to compare the outcomes of models with two and three classes. The architecture chosen for the different tests impacts the model's input. Investigations have been conducted on 2D (slice-level) and 3D (subject-level) methodologies. A 2D solution is typically used. The benefit that enables a larger collection of instances is provided by the 176 slices that make up each MRI volume. As a result, training convergence may be improved, and overfitting may be decreased. However, the third axis of 2D techniques needs to be improved in information. When working with 3D things, this knowledge could offer specifics that lead to a more precise solution, such as human tissues and organs. Both well-known and unique architectures are considered for 2D and 3D techniques. ResNet18 gave the greatest results when the residual network (ResNet) family was evaluated against inception, Xception, and other well-known architectures. Both 2D and 3D convolution techniques have been applied to construct custom networks. Next, the network architectures that have been developed and compared are described.

3.3. 2D convolution

The proposed custom network is a small size. A single output is considered available for the entire volume, with the input data being at the subject level of 3D images. Here, M , N , and K comprise the input data, where K specifies total slices. Each slice sizes are M and N . Slice-by-slice processing of the data using reduced spatial resolution is achieved by using 2D convolutional layers and maximum pooling layers, which produce a sample of high-level descriptors typical of the data. 2D max pooling layer and the 2D convolutional layer with 3 filter sizes are present in each of the four convolutional blocks. There are no extra completely connected layers introduced. "Softmax" refers to the activation mechanism for exclusive classes. Categorical cross-entropy is the loss function. The model shows the addition of batch normalization layers, an enhancement method. Working with 3D images can provide information that will help in providing a more accurate response, just like it does with human tissues and organs. Both well-known and atypical architectures are considered for 2D and 3D techniques. With inception, Xception, ResNet, and other well-known architectures, ResNet18 produced the best outcomes. Both 2D and 3D convolution techniques have been applied to construct custom networks. The developed and compared network architectures are then described. A simple custom network is suggested using 2D conv. A single output is considered available for the entire volume, with the input data being at the level of a 3D topic. Here, M , N , and K make up the input data, where K specifies total slices. Slice-by-slice processing of the data using reduced spatial resolution is achieved by using 2D convolutional layers and maximum pooling layers, which also produce a sample of high-level descriptors of typical data. The 2D max pooling layer and the 2D convolutional layer with 3 filter sizes are present in each of the four convolutional blocks. There are no extra completely connected layers introduced. "Softmax" refers to the activation mechanism for exclusive classes. Categorical cross-entropy is the initial loss function. Comparing two iterations of the same design, Figure 2 depicts the fundamental structure of the 2D system. It shows the addition of batch normalization layers as an enhancement method.

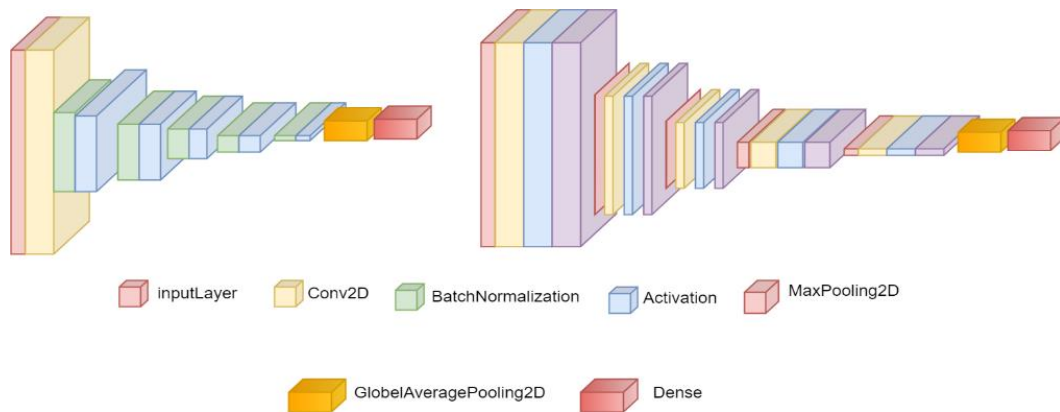


Figure 2. Fundamental 2D system

3.4. 3D model

It is suggested to use a custom 3D network with a true 3D approach similar to the prior one. Here, M, N, K , and 1 are the input data, with K indicating the total number of slices and the letters M and N denoting the size of a slice. Utilizing five convolutional blocks, each consisting of a 3D max pooling layer and a 3D convolutional layer, is advised. Because it is more in tune with biology and human cognition, this technique may produce more favorable patterns of SCI creation. The filters specify total filters and filter size for every convolutional layer. Five convolutional blocks contain 8, 16, 32, 64, and 128 filters, with an average filter size of 3. The 3D global average pooling layer is applied after the embedding phase. There are only a few new, fully connected layers added. "Softmax" refers to the activation mechanism for exclusive classes. Categorical cross-entropy serves as the first loss function. It displays two versions of the same architecture.

The Net2D network employed batch normalization. An original 3D network that employs a true 3D method is suggested. It is comparable to the prior network. Here, M, N, K , and 1 are the input data, with K indicating the total number of slices and the slice's size indicated by M and N . It is recommended to utilize five convolutional blocks, each comprising a 3D convolutional layer and a 3D max pooling layer. This strategy may result in more advantageous patterns of brain formation because it is more in line with biology and human cognition. The filters specify number of filters and * specifies filter size for every convolutional layer. Five convolutional blocks contain 8, 16, 32, 64, and 128 filters with an average filter size of 3. The 3D global average pooling layer is applied after the embedding phase. There aren't any more completely connected levels. "Softmax" refers to the activation mechanism for privileged classes. Categorical cross entropy serves as the first loss function. Figure 3 depicts the 3D architecture. We compare two iterations of the same design once again. The baseline Net3D architecture and the batch normalization layers are added as a tool for enhancement.

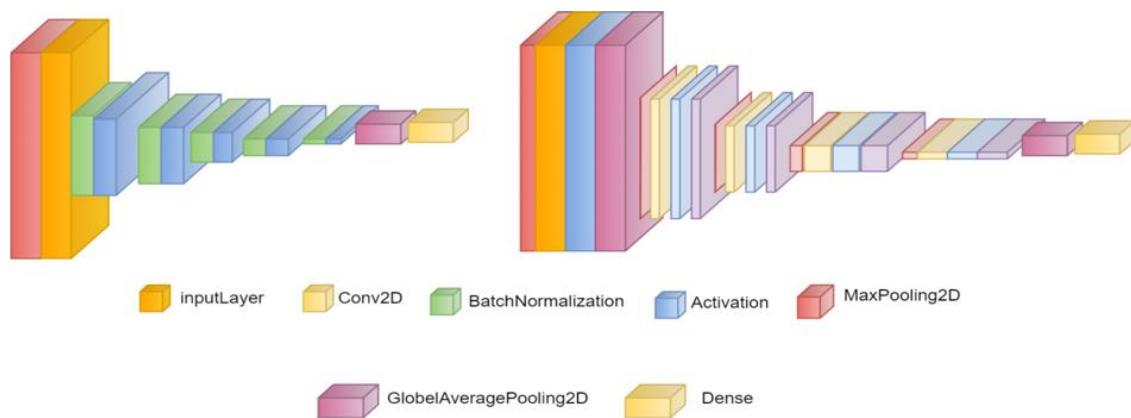


Figure 3. Proposed 3D network model

3.5. Slice network

The network architecture's performance is comparable to 2D slices. It implies that each image has a corresponding categorization output and that the network only receives a single slice of M, N , and 3 sizes. The three-channel input needed for ResNet family networks was produced using a single separate channel per slice duplicated, enabling ImageNet weights to fine-tune processes. Reference models were created using a tiny ResNet family model honed using ImageNet. Similar to the earlier networks, after the embedding phase, the loss function is "categorical_crossentropy," while the activation function is "softmax" It is important to remember that the final result of a full study is expressed as a "majority vote" across all sequence slices. People used to DL-based innovations are aware that the dataset and the specifics of the problem have some problems. The outcomes might be enhanced if best practices are considered. The following have been the more notable tactics used: the learning rate enables an iterative oscillation between two values for the learning rate. Conventional, well-known instructional techniques from the past typically consider gradually slowing down learning as time passes and utilizing various functions (linear, polynomial, and step). This tactic may cause the model to sink into regions with low loss values. With learning rate, it is simple to identify the ideal learning rate parameter, allowing for improved (and quicker) model tuning. After being defined, the learning rate will fluctuate between a minimum and a maximum value. It is possible to define the various working policies: simple triangular cycles include "triangular," "triangular2," and "exp range," which are all triangular cycles with the addition that in every cycle, the maximum learning rate is reduced by half.

A method for training deep neural networks called batch normalization standardizes each mini-input batch to a layer before it enters the network. As a result, the learning process is stabilized, frequently improving training process and model performance. When a dataset is too small, one of the key problems is over-fitting. Batch normalization can help to mitigate this problem. The 3D-subject level technique is particularly notable for this over-fitting issue. BatchNormalization () method and a "rectified linear unit (ReLU)" activation are used in batch normalization, which comes before and after max pooling layer and convolutional layer. The proposed network model uses processing layers for reducing the input images to the related key features so that the provided samples can be easily categorized. The benefits of the proposed model over other classification approaches are its ability to learn the key characteristics of its own and reduce the necessity of hyper-parameters and filtering. Also, the model provides better prediction outcomes with the provided input.

4. DISCUSSION

There were determined to be 289 individuals with MRI scans that may be used to build models. The 201 patients in the training/validation dataset had 6,588 training images. The holdout dataset contained 2,991 individual images of 88 patients. Each dataset's demographic data are displayed. The provided dataset did not significantly differ in patient variables like age and gender, nor did baseline mJOA, the manufacturer of the MRI scanner, or the MRI image parameters. Holdout dataset holds 2,991 images and training/validation dataset holds 6,588 images were examined by two independent raters who looked for images that showed signs of spinal cord compression that is circumferential or partial. On the training/validation dataset, concordance between the two raters shows Cohen's $\kappa = 0.83$ and $\kappa = 0.83$.

The online available dataset served as the basis for model's training. The collection included 6,588 distinct axial images from 201 patients. A training cohort (4,941 images) and a validation cohort (201 patients) were created using the 201 patients' axial images and labels (1,647 images). In this study of learned models, the binary cross-entropy loss did not decrease after training seven distinct neural network topologies for 10 epochs. Model 6 was retained for additional testing due to its improvement on the validation dataset; accurate classification and minimal binary cross-entropy loss were accomplished. The ResNet-50 CNN in Model 6 included two entirely coupled layers that each included 512 neurons, two dropout levels with a 30% dropout, and two fully connected layers.

The model generated 94% AUC, 88% sensitivity, 89% specificity. The classification made by humans in the holdout dataset was compared to each slice's deep-learning classification output. ROC curve and an AUC measurement were produced by contrasting each slice's expected and actual classes of all patients. Holdout dataset's area under curve (AUC) was 0.94 as in Figure 4. Each sub-group AUC is evaluated for measuring performance. AUC did not substantially fluctuate from AUC of holdout dataset based on $p - values > 0.05$ as in Figures 5 and 6. Figure 4 displays activation maps for fewer samples of model's true positive (TP) and false negative (FN) predictions. To understand the traits that indicated whether or not a class was compressed, we looked at these images. TP classifications were considered for SCI and corresponding spaces. Two clinically sample regions of class activation maps (CAMs) looked to be active. The model occasionally appears to rely on elements encompassing the para-spinal muscles and vascular

systems outside the spinal canal in the sample photographs of inaccurate negative classifications. The spinal cord had clinical importance, which appeared to be the algorithm's primary emphasis among the false negative images. Nevertheless, this led to an inaccurate categorization.

This work used a dataset of 289 DCM patients to create and evaluate MRI imaging of the cervical spine and used it to determine spinal cord compression using the 2D-CNN model. We gave evidence that a unique medical imaging classification challenge may be accomplished by training an existing CNN model. In our holdout dataset, the model performed well across a range of patient classifications and scanner types (94% AUC). Our study is the first to develop a model for identifying spinal cord compression using a sizable prospective dataset of highly accurate cervical spine MRI scans. This methodology may be helpful in a clinical trial setting to enable quick automatic coding of cervical spine MRI data. This approach may enhance the impartiality and accuracy of reading MRI data for the cervical spine. The importance of ML methods using 2D-CNNs for future medical diagnoses is becoming more widely acknowledged. The automated interpretation of imaging in spinal diseases using CNNs has to be more thoroughly established, despite systems applying analogous methodologies having been created and given regulatory permission for chest radiography, mammography, and imaging of brain damage. Most studies similar to ours have concentrated on categorizing lumbar spine MRI images. Using available MRI scans, CNN is used to categorize disc narrowing, lumbar disc degeneration, upper/lower marrow changes, upper/lower endplate abnormalities, spondylolisthesis, and central canal stenosis. Table 1 describes the patient characteristics. The AUC and p-value of the proposed model is described. The hold-out range is measured for < 40, 40–65 and >65. The age ranges from 88% to 95% and p-value ranges from 0.1 to 0.4.

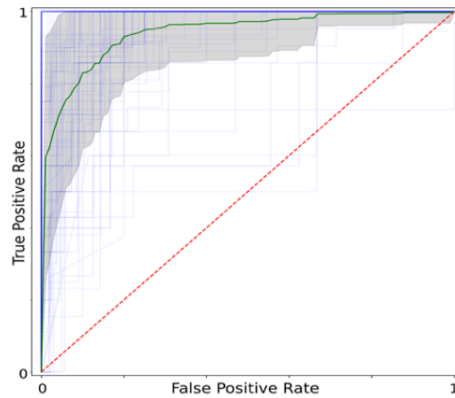


Figure 4. AUC

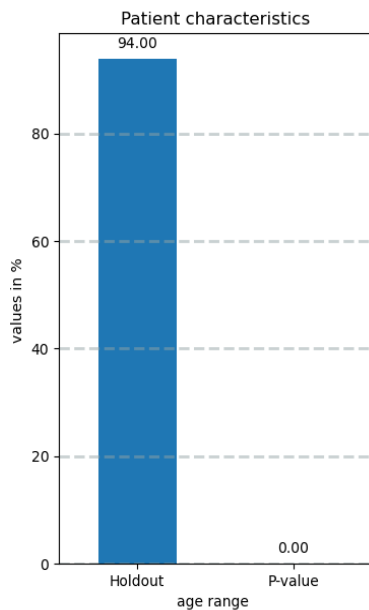


Figure 5. Hold-out and p-value comparison

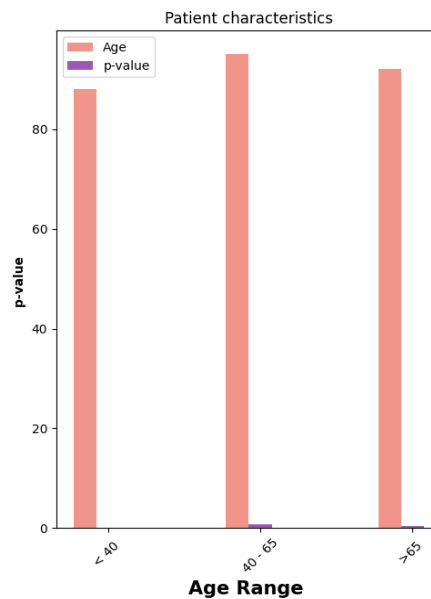


Figure 6. P-value analysis

Table 1. Patient characteristics

	AUC (%)	p-value
Holdout	94	--
	Age (%)	
< 40	88	0.1
40 – 65	95	0.7
> 65	92	0.4

Visual geometry group (VGG) CNN was used by these researchers, who trained the model using randomly initialized weights. The deep network architecture created a significant dataset. This study created 2D-CNN with an accuracy of 84.5% for evaluating foraminal and lumbar canal stenosis. The intervertebral disks and vertebral bodies were divided up using a segmentation approach used by these authors, and the ResNet-50 CNN was used to classify the output of this model. T2-weighted images of spine were used to train the ResNet-50 CNN instead of going through a separate segmentation stage. Image segmentation allows it to automatically locate pathologic results, which was impossible with the more straightforward approach we employed. A model is developed to assess canal stenosis and lumbar disc herniation; 3,560 patients' lumbar spine MRI scan data were employed. Intervertebral disks were split during a segmentation step in their procedure, which is quite similar to the deep spine architecture. The next stage involved a specially designed 2D-CNN in evaluating the degree of disk herniation and channel stenosis. The AUC for the disk herniation detection by this study's authors was 0.808. Our approach was distinct from this one in that we used a binary classification technique to distinguish between compressed and uncompressed spinal cords rather than rating the severity of spinal cord compression.

The lumbar spine has been the subject of earlier research on the automated analysis of degenerative spine images. However, several published publications have used deep-learning models to analyze neck or cervical vertebra images. CNN was trained to identify fatty infiltration in cervical spine MRI data to gauge the degree of fatty infiltration. They then showed how the model's output parameters matched clinical metrics, including neck discomfort and neck-related disability. Imaging dataset includes DTI analyses of individuals who later under-went surgery. Table 2 depicts the comparison of training and validation process as shown in Figures 7 and 8. The accuracy of the ML model this researcher built to predict a successful surgical result in their training cohort was 91.1% as shown in Figures 9 to 12. Due to the small size of their cohort of 35 patients-they could not fully validate their model. Two labelers looked at each image as part of our data labeling process and assigned either a compressed or uncompressed label. The two labelers' agreement was high on both datasets with Cohen's Kappa's of 82% and 83%, respectively.

Table 2. Training and validation process

Dataset	Training and validation		Hold-out	
	Compress (%)	Non-compress (%)	Compress (%)	Non-compress (%)
Label 1	24	77	21	79
Label 2	21	80	20	81

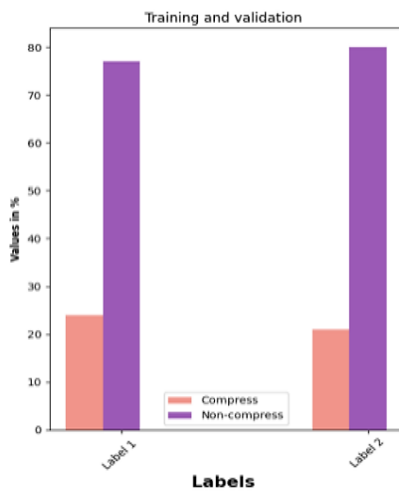


Figure 7. Training and validation-based compress and non-compress

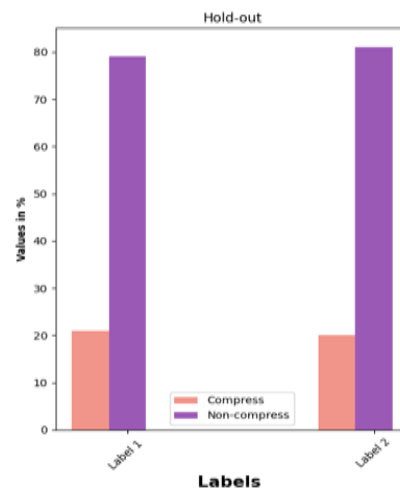


Figure 8. Hold-based compress and non-compress

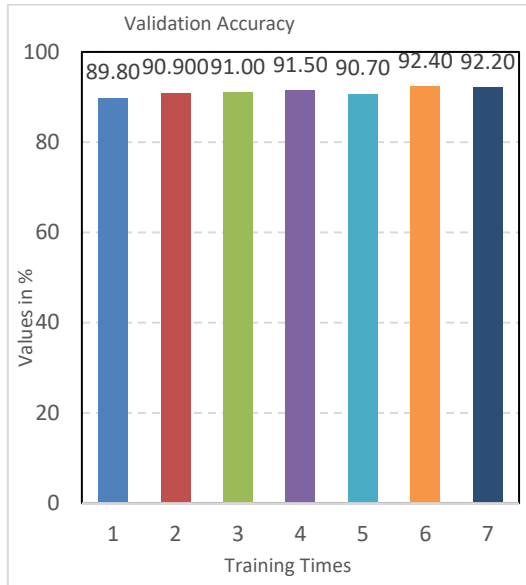


Figure 9. Validation accuracy

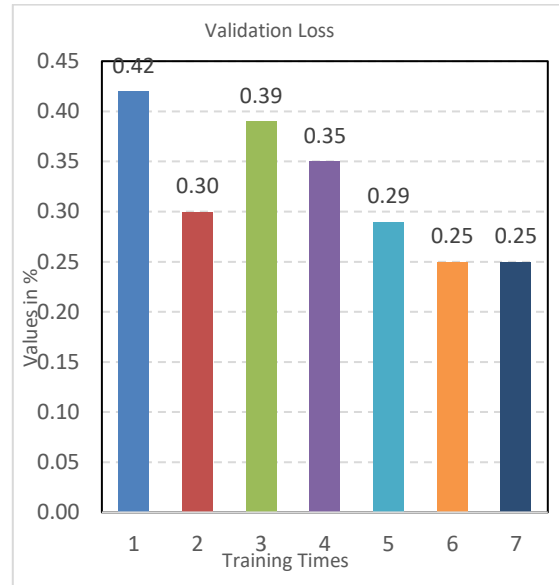


Figure 10. Validation loss

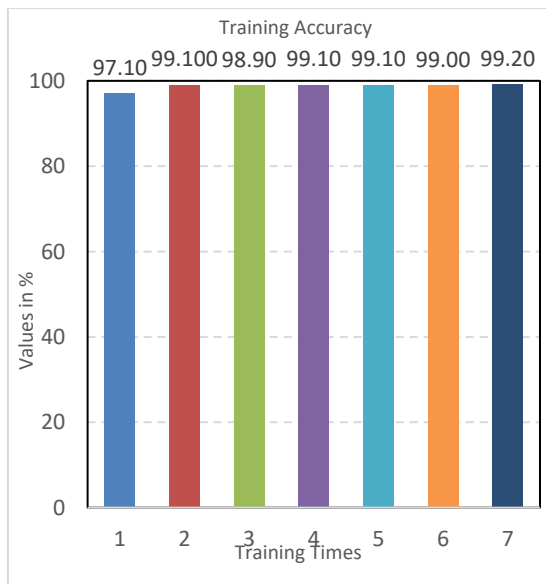


Figure 11. Training accuracy



Figure 12. Training loss

Table 2 gives a training and validation with compress and non-compress. For label 1 and label 2, the compress is 24% and 21% and non-compress is 77% and 80% (training and validation). For label 1 and label 2, the compress is 21% and 20% and non-compress is 79% and 81% (Hold-out). Table 3 describes the experimental outcomes of the proposed model for 50, 55, 60, 70, and 75 epochs respectively. The accuracy and loss during training and validation is provided. The proposed model shows an accuracy of 89.8% to 92.4% accuracy during validation process and 97.1% to 99.1% accuracy during training process. Similarly, the loss during validation is 0.25 to 0.42 and loss during training process is 0.021 to 0.089 respectively.

Although objective standards have been discussed, clinical evaluation is frequently subjective, making it challenging to gauge SCI. Images where the two raters disagreed usually showed modest degenerative alterations but no appreciable SCI. Finally, labeling was chosen. Instead of using a binary classification as we did, future studies in this field may classify the cervical spinal cord compression level and try to create a model. It can be achieved by labeling cervical spinal MRIs using a grading scheme. This approach yields subjective scores for four different types of SCI and internal rate of return (IRR) produced 0.5 Cohen's Kappa. We achieved superior with simpler binary grading system.

Table 3. Performance comparison during training

Training	Epochs	Validation		Training	
		Accuracy (%)	Loss	Accuracy (%)	Loss
1	50	89.8	0.42	97.1	0.089
2	50	90.9	0.30	99.1	0.021
3	50	91	0.39	98.9	0.029
4	55	91.5	0.35	99.1	0.026
5	60	90.7	0.29	99.1	0.024
6	75	92.4	0.25	99	0.028
7	70	92.2	0.25	99.2	0.023

We experimented with different model architectures when training the models. All models were employed using the ResNet50 CNN architecture, which had fully linked layers that varied amongst the models and pre-trained weights. We discovered all of the evaluated models as shown in Table 3, whose accuracy ranged from 89.88% to 92.39%, performed admirably. Two layers with 512 neurons each made up the model with the best performance that was fully linked, and two layers, each with 30% dropout. The holdout dataset showed good performance for this model, with 94% AUC and p-value comparison is shown in Figures 4 and 5. A successful lumbar spine models developed have performance parameters equivalent to those utilized in this investigation. We used test images that were accurately categorized from the testing dataset (true positives) and wrongly classified (false positives) to create CAMs (false negatives) as in Figures 7 and 8. There are problems with class activation maps, and they may experience issues with reproducibility or repeatability. CAMs have been demonstrated to be particularly ineffective at locating pathogenic characteristics. In this case, localization or segmentation models are suggested for medical imaging data. CAMs still only provide a marginal advantage, even when utilized to emphasize characteristics that a model thought were crucial for prediction during a post-hoc analysis. When labels were randomly permuted, they occasionally revealed visually convincing portions of the image. Based on these restrictions, we interpret the outcomes of our CAMs. The clinically relevant areas of the spinal cord and cerebrospinal fluid (CSF) space were highlighted in the true positive images when we looked at CAMs for a couple of sample images. In the false negative photographs, which made up a relatively minor fraction of the classifications, the CAMs frequently suggested activity over irrelevant image regions, such as the vascular systems or paraspinal muscles. It could imply that the model is giving more weight to the spinal cord compression features in the image rather than any other less significant abnormalities that might be present. However, this would need to be confirmed rather than only looking at a portion of the testing photographs. Although this is outside the purview of the current study, more investigation should be done into the CAMs' consistency, repeatability, and sensitivity to model weight randomization.

Our research has a few drawbacks. First, only patients with a confirmed diagnosis of SCI who had surgery were included in our dataset. As a result, a small percentage of the patients in our sample had mild SCI or normal MRI results. A more diversified training image set and generalizable model is developed by integrating asymptomatic patients with mild symptoms throughout model's training. Because a significant fraction of the patients required MRI in the correct format, the total images used for model training needed to be increased. Two data labelers were utilized and we used a consensus mechanism to resolve discrepancies. Our set of ground truth labels would have been of higher quality with accurate assessment of model performance would have been possible. This work did not differentiate circumferential and partial SCI during data labeling and model training. Despite being more dangerous in clinical symptoms, circumferential SCI is a common clinical practice outcome.

A classification scheme for cervical spinal cord compression of more than two groups may be created as a consequence of future research in this area. This model would be more useful therapeutically. MRI data from the cervical spine can be used to identify spinal cord compression as shown in Figures 6 and 7. We developed 2D-CNN. Our model might be helpful for clinical studies that automatically code MRI scans as they stand now. The model might be applied in this scenario to automatically extract properties from MRI images recorded as part of a clinical trial, like the percentage of slices exhibiting spinal cord compression, which could provide information for follow-up research. We acknowledge that issues still need to be resolved before our model can be applied in a medical environment. Our model does a certain type of classification. A complete model will enable extremely precise diagnosis and distinction of a wide range of pathologies, including foraminal stenosis and disc herniation; some symptoms include spondylolisthesis, ligamentous hypertrophy, and posterior longitudinal ligament ossification. To build and test such a model, it is probably required to have a dataset with thousands of cases. According to other researchers, large imaging datasets are a major obstacle to developing ML algorithms with therapeutic relevance. Our findings imply that 2D-CNN can be trained in spine surgery for a unique medical imaging classification task. However, there are still issues to be resolved. A larger training dataset and a more general model might both be

produced using our strategy. From Table 4 it is obvious that the proposed model works well compared to various existing approaches and gives promising solution.

Table 4. Overall performance comparison

Methods	Accuracy (%)	Precision (%)	Recall (%)
PCA ensemble	66	68.64	60.25
PSO ensemble	69	69.35	62.33
SVM + linear kernel	72	73.84	64.65
MMD	75	75	66.38
Proposed model	92	91	92

5. CONCLUSION

Automated diagnostic technology of SCI has made significant strides in recent years, and this trend is predicted to continue. Here, a novel 2D-CNN model was trained and put through its paces using spine MRI data to determine spinal cord compression. With an AUC of 0.94, we successfully produced a model that performed well across a large patient group. We demonstrated how an existing standard CNN might be trained to complete a challenging medical imaging classification assignment. To evaluate the SCI severity, cervical spine deformity, cord signal alteration and nerve root compression, future research will need to concentrate on creating larger datasets. Improved radiology operations and clinical decision-making could result from increased objectivity and efficacy in cervical spine MRI interpretation.




REFERENCES

- [1] M. L. Giger, "Machine Learning in Medical Imaging," *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512–520, Mar. 2018, doi: 10.1016/j.jacr.2017.12.028.
- [2] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 86–101, May 2019, doi: 10.1016/j.zemedi.2018.12.003.
- [3] W. Li, J. Xu, X. Chen, J. He, and Y. Huang, "Phase Synchronization Between Motor Cortices During Gait Movement in Patients with Spinal Cord Injury," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 1, pp. 151–157, Jan. 2016, doi: 10.1109/tnsre.2015.2453311.
- [4] S. J. Qin and L. H. Chiang, "Advances and opportunities in machine learning for process data analytics," *Computers & Chemical Engineering*, vol. 126, pp. 465–473, Jul. 2019, doi: 10.1016/j.compchemeng.2019.04.003.
- [5] I. Seanez-Gonzalez et al., "Static Versus Dynamic Decoding Algorithms in a Non-Invasive Body–Machine Interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 7, pp. 893–905, Jul. 2017, doi: 10.1109/tnsre.2016.2640360.
- [6] S. K. H. Ahammad, V. Rajesh, and M. D. Z. U. Rahman, "Fast and Accurate Feature Extraction-Based Segmentation Framework for Spinal Cord Injury Severity Classification," *IEEE Access*, vol. 7, pp. 46092–46103, 2019, doi: 10.1109/access.2019.2909583.
- [7] P. Sok, T. Xiao, Y. Azeze, A. Jayaraman, and M. V. Albert, "Activity Recognition for Incomplete Spinal Cord Injury Subjects Using Hidden Markov Models," *IEEE Sensors Journal*, vol. 18, no. 15, pp. 6369–6374, Aug. 2018, doi: 10.1109/jsen.2018.2845749.
- [8] A. A. Reshi, I. Ashraf, F. Rustam, H. F. Shahzad, A. Mehmood, and G. S. Choi, "Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms," *PeerJ Computer Science*, vol. 7, pp. 1–34, Jul. 2021, doi: 10.7717/peerj-cs.547.
- [9] J. Alabort-i-Medina and S. Zafeiriou, "A Unified Framework for Compositional Fitting of Active Appearance Models," *International Journal of Computer Vision*, vol. 121, no. 1, pp. 26–64, Jun. 2016, doi: 10.1007/s11263-016-0916-3.
- [10] K. L. Khatri and L. S. Tamil, "Early Detection of Peak Demand Days of Chronic Respiratory Diseases Emergency Department Visits Using Artificial Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 285–290, Jan. 2018, doi: 10.1109/jbhi.2017.2698418.
- [11] S. H. Hojjati, A. Ebrahimzadeh, A. Khazae, and A. Babajani-Feremi, "Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM," *Journal of Neuroscience Methods*, vol. 282, pp. 69–80, Apr. 2017, doi: 10.1016/j.jneumeth.2017.03.006.
- [12] B. Sánchez et al., "Machine learning on difference image analysis: A comparison of methods for transient detection," *Astronomy and Computing*, vol. 28, pp. 1–17, Jul. 2019, doi: 10.1016/j.ascom.2019.05.002.
- [13] Y. Shi, P. Li, H. Yuan, J. Miao, and L. Niu, "Fast kernel extreme learning machine for ordinal regression," *Knowledge-Based Systems*, vol. 177, pp. 44–54, Aug. 2019, doi: 10.1016/j.knosys.2019.04.003.
- [14] P. Sherubha, S. P. Sasirekha, V. Manikandan, K. Gowsic, and N. Mohanasundaram, "Graph based event measurement for analyzing distributed anomalies in sensor networks," *Sādhanā*, vol. 45, no. 1, pp. 1–5, Dec. 2020, doi: 10.1007/s12046-020-01451-w.
- [15] P. Sherubha and N. Mohanasundaram, "An Efficient Network Threat Detection and Classification Method using Anp-Mvps Algorithm in Wireless Sensor Networks," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 11, pp. 1597–1606, Sep. 2019, doi: 10.35940/ijitee.k3958.0981119.
- [16] K. U. K. M., "Energy Efficient Method of Detection and Prevention of Wormhole Attack in Wireless Sensor Networks Using NS-2 Simulator," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 12, no. 1, pp. 775–779, Feb. 2020, doi: 10.5373/jardcs/v12sp1/20201128.
- [17] C. S. Ahuja et al., "Traumatic Spinal Cord Injury—Repair and Regeneration," *Neurosurgery*, vol. 80, no. 3, pp. 9–22, Feb. 2017, doi: 10.1093/neuros/nyw080.
- [18] S. S. Panesar, R. N. D'Souza, F.-C. Yeh, and J. C. Fernandez-Miranda, "Machine learning versus logistic regression methods for




- 2-year mortality prognostication in a small, heterogeneous glioma database,” *World Neurosurgery: X*, vol. 2, pp. 1-12, Nov. 2018, doi: 10.1101/472555.
- [19] S.-I. Lee *et al.*, “A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia,” *Nature Communications*, vol. 9, no. 1, pp. 1-13, Jan. 2018, doi: 10.1038/s41467-017-02465-5.
- [20] Y. Shen, Y. Li, H.-T. Zheng, B. Tang, and M. Yang, “Enhancing ontology-driven diagnostic reasoning with a symptom-dependency-aware Naïve Bayes classifier,” *BMC Bioinformatics*, vol. 20, no. 1, pp. 1–14, Dec. 2019, doi: 10.1186/s12859-019-2924-0.
- [21] D. B. McCoy *et al.*, “Convolutional Neural Network–Based Automated Segmentation of the Spinal Cord and Contusion Injury: Deep Learning Biomarker Correlates of Motor Impairment in Acute Spinal Cord Injury,” *American Journal of Neuroradiology*, vol. 40, no. 4, pp. 737–744, Mar. 2019, doi: 10.3174/ajnr.A6020.
- [22] A. V. Karhade *et al.*, “Machine learning for prediction of sustained opioid prescription after anterior cervical discectomy and fusion,” *The Spine Journal*, vol. 19, no. 6, pp. 976–983, Jun. 2019, doi: 10.1016/j.spinee.2019.01.009.
- [23] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
- [24] J. Wen *et al.*, “Convolutional neural networks for classification of Alzheimer’s disease: Overview and reproducible evaluation,” *Medical Image Analysis*, vol. 63, pp. 1-39, Jul. 2020, doi: 10.1016/j.media.2020.101694.
- [25] V. Venkatraghavan, E. E. Bron, W. J. Niessen, and S. Klein, “Disease progression timeline estimation for Alzheimer’s disease using discriminative event-based modeling,” *NeuroImage*, vol. 186, pp. 518–532, Feb. 2019, doi: 10.1016/j.neuroimage.2018.11.024.
- [26] A. V. Karhade *et al.*, “Machine learning for prediction of sustained opioid prescription after anterior cervical discectomy and fusion,” *Spine Journal*, vol. 19, no. 6, pp. 976–983, 2019, doi: 10.1016/j.spinee.2019.01.009.
- [27] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.
- [28] J. Wen *et al.*, “Convolutional neural networks for classification of Alzheimer’s disease: Overview and reproducible evaluation,” *Medical Image Analysis*, vol. 63, 2020, doi: 10.1016/j.media.2020.101694.
- [29] S. Vieira, W. H. L. Pinaya, and A. Mechelli, “Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications,” *Neuroscience and Biobehavioral Reviews*, vol. 74, pp. 58–75, 2017, doi: 10.1016/j.neubiorev.2017.01.002.
- [30] V. Venkatraghavan, E. E. Bron, W. J. Niessen, and S. Klein, “Disease progression timeline estimation for Alzheimer’s disease using discriminative event based modeling,” *NeuroImage*, vol. 186, pp. 518–532, 2019, doi: 10.1016/j.neuroimage.2018.11.024.

BIOGRAPHIES OF AUTHORS






P. R. S. S. Venkatapathi Raju    received M.Tech. degree (Computer Science and Technology) in 2011 from JNTU Kakinada University. He has published number of papers in different International Conferences and Journals. He has 10+ years teaching experience. His research interests are including image processing, data mining, machine learning, and deep learning. He is member of CSI (I1501635). He can be contacted at email: prssvraju2@gmail.com.



Valayapathy Asanambigai    received M.E degree in CSE from Annamalai University. She completed his Ph.D. degree in Computer Science and Engineering at Annamalai University. She is currently an Assistant Professor in Computer Science and Engineering at Annamalai University since. She is having 15+ teaching experience. She published more than 12 papers in international journals and conferences. His research interests are including image processing, data mining, and machine learning. She can be contacted at email: tradingbaskeran@gmail.com.



Suresh Babu Mudunuri    completed his Masters in IT from IIIT Hyderabad in 2005 and his PhD in Computer Science & Systems Engineering from Andhra University College of Engineering in 2012. He is currently working as Head Centre for Bioinformatics Research at SRKR Engineering College, Bhimavaram and has published 20+ research papers in reputed journals and international conferences. He received the Young Researcher Award from IBCB in the year 2016 and has successfully completed a DST Funded Project worth 22 Lakh. He can be contacted at email: sureshmudunuri@gmail.com.