Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 5, October 2025, pp. 4001~4012
ISSN: 2302-9285, DOI: 10.11591/eei.v14i5.10617

Hybrid DL and ML approach for MRI-based classification of bone marrow changes in lumbar vertebrae

Yasir Hussein Shakir[1], Tiong Sieh Kiong[2], Chai Phing Chen[3], Sachin Sharma Ashok Kumar[4]
[1]Department of Engineering, College of Graduate Studies (COGS), Universiti Tenaga Nasional (UNITEN), Kajang, Malaysia
[2]Institute of Sustainable Energy (ISE), College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, Malaysia
[3]Department of Electrical and Electronics Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, Malaysia
[4]School of Engineering, Faculty of Innovation and Technology, Taylor's University, Subang Jaya, Malaysia

Article Info

Article history:

Received May 4, 2025
Revised Aug 29 2025
Accepted Sep 11, 2025

Keywords:

Bone marrow changes lumbar vertebrae
Classification
ConvNeXt
Hybrid mode
Machine learning

ABSTRACT

Alterations in the bone marrow changes lumbar vertebrae (BMCLVB) are considered important markers of chronic low back pain severity, particularly among patients with coexisting conditions like osteoporosis or cancer. Realizing these associations informs healthcare and insurance frameworks but also supports early intrusion planning for high-risk populations. This study aims to classification (BMCLVB) as normal or abnormal used magnetic resonance imaging (MRI) with machine learning (ML) model. A novel dataset comprising 1,018 BMCLVB MRI images was utilized to extract deep features via a pre-trained ConvNeXt-XLarge model. These features were then classified using different types in individual and ensemble ML algorithms. To ensure a comprehensive performance evaluation, all models were tested using accuracy, precision, recall, and F1-score. The combination of ConvNeXt-XLarge and logistic regression (LR) achieved a classification accuracy 93.14%, precision 93.22%, recall 94.83%, and F1-score 94.02%. These results highlight that the proposed model provides a fast and cost-efficient solution supporting the diagnosis of BMCLVB and potential to significantly improve clinical decision-making and patient care outcomes.

Corresponding Author:

Tiong Sieh Kiong
Institute of Sustainable Energy (ISE), College of Engineering, Universiti Tenaga Nasional (UNITEN)
Kajang, Malaysia
Email: siehkiong@uniten.edu.my

1. INTRODUCTION

Magnetic resonance imaging (MRI) of spine is frequently performed patients with low back pain because it allows direct visualization of the vertebral column, spinal cord and nerves, and adjacent soft tissues. Disc abnormalities such as bulges or herniations ability compress surrounding nerves, may result in paresthesia of the legs and on advanced cases, neuropathic injury accompanied by lower limb weakness. MRI provides without surgery evaluate bone marrow through detection of shifts in water and fat levels, which are indicative of pathological changes because of its superior sensitivity to fat content and comparatively minor influence from other factors mineralized tissue on the MRI signal [1]. Initial applications of artificial intelligence (AI) in spinal evaluations have shown significant value in identifying localized abnormalities. For example, some models have been known to detect early signs associated with compressive myelopathy and demyelinating spinal lesions not easily seen on regular MRI for clinicians or researchers less familiar with AI, some terminology in this field may be unfamiliar. However machine learning (ML) and deep learning (DL) approaches explain distinct concepts, both were often erroneously apply interchangeably [2].

This can result from mistakes about specific capabilities and applications of these technologies in medical imaging [3]. There are available AI algorithms in spine imaging that are important for picture quality enhancement evaluation, automatic labeling of vertebral level, spine lesion localization, as well as segmentation and characterization [4]. There is growing evidence pointing to the utilization on techniques for comprehending nature of the spine lesion, as well as indicating that the picture characteristics that may be useful for both lesion identification and classification have solid potential to be identified by these methods. There are two most commonly utilized ML approaches, which consist of feature analysis with radiomics and convolutional neural networks (CNNs), respectively. Radiomics entails deriving several predefined facets for developing related datasets used in training DL algorithms for image categorization [5]. However, as for the disadvantage, this method is fully based on the allowance of the radiologist or clinician for selecting features, and this severely affects the precision of the algorithm [6]. Consequently, machine and DL strategies are capable of identifying pertinent imaging features relevant to classification on their own and there is no need for the feature selection process. This typically entails the use of CNNs, which the studies show improve the accuracy in identification of lesions, their delineation, and assessment of tumor responses to treatment in cancer imaging [7]-[9]. The classification of bone marrow alterations in lumbar vertebrae (BMCLVB) is a decisive but complex assignment through the areas of attention in medical imaging. Therefore, the purpose in this review is to supply a concise and focused overview of the ML applications in spinal oncology imaging, which include radiomics and CNN. Furthermore, the development and evaluation of models for this task remain challenging, largely due to the absence of a dedicated in public available dataset. Creating this dataset signifies a major milestone in employing DL approaches techniques to (BMCLVB).

Our research makes the following novel contributions:

– We collected and curated a unique dataset comprising 1,018 MRI images from 134 patients, specifically targeting BMCLVB. To our knowledge, this is the first dedicated dataset of this scale for BMCLVB classification.
– Unlike previous studies that first applied the CNNs or radiomics-based feature engineering, we leverage the state-of-the-art ConvNeXt XLarge architecture to automatically extract 4,096-dimensional feature vectors and hierarchical representations of lumbar bone marrow patterns.
– We experiment evaluated nine ML classifiers involving K-nearest neighbors (KNN), support vector machine (SVM), decision trees (DT), multiple perceptron (MLP), random forest (RF), histogram-based gradient boosting (HGB), adaptive boosting (AdaBoost), passive aggressive (PA), and logistic regression (LR), extracted features and this inclusive comparison demonstrates the performance of combining ConvNeXt features with classical ML demonstrated several existing approach in terms of accuracy and robustness.
– The strengths and weaknesses of (BMCLVB) classifiers are best revealed through a comprehensive evaluation based on accuracy, precision, recall, F1-score, and confusion matrices.

2. PREVIOUS WORKS

The implementation for ML/DL approaches in spinal imaging have developed in recent years with promising results in areas such as disc degeneration, spondylolisthesis, and spinal cord compression. Nevertheless, a significant gap in automated diagnostic evidence BMCLVB leaving a critical gap in automated diagnostic support to detect diagnosis. In the field of radiomics analysis to predict the risk of bone metastasis in various studies [10]. Lim *et al.* [11] evaluated whether this approach may improve radiologists' performance in follow-up research. Eight radiologists' performance was assessed by the authors both and without DL model support. They discovered that DL model help resulted in considerable time reductions (76–203 s and $p<0.001$), with in-training radiologists benefiting the most. When compared to the baseline, readers who received DL assistance performed better or similarly. This study by Xie *et al.* [12] applied a merger models optimized MedSAM a radiomics-enhanced framework for automatic Pfirrmann classification of cervical disc degeneration. Trained on sagittal T1 and T2-weighted images, the model yielded an area under the receiver operating characteristic (ROC) curve of 95.00%, an accuracy of 89.51%, precision of 87.07%, recall of 98.83%, and F1-score of 93.00% in the test evaluation. Lin *et al.* [13] proposed an attention U-Net framework, defining spinal inflammation by the presence of active inflammatory lesions on the STIR sequence. The model was tested on MRI images and gave an area under the curve (AUC) of 87.00%, a sensitivity of 80.00%, and a specificity of 88.00%. The deep neural network offers the potential to broaden the use of spinal MRI for treating axSpA. The extra trees classify is the most effective for classification in this study [14], comprising a diverse set of ML models such as DT models, SVMs, and neural networks. The models show that ML models are successful in identifying IVD disease, with accuracy and precision reaching up to 90.83% and 91.86%, respectively. Trinh *et al.* [15] developed computer-aided diagnostic (CAD) called LumbarNet assesses the effectiveness of model in automatic identifying, spondylolisthesis from lumbar x-ray. The model's accuracy in detecting vertebral slip was 88.83%. We find that LumbarNet performed better than U-Net, a popular technique for segmenting medical images, and may be a trustworthy way to detect spondylolisthesis. The DL-based technique capable of detecting cervical spinal cord compression using MRI data was developed and validated in this work [16], the resulted in an overall achieving 94.00% AUC, 88.00% sensitivity, 89.00% specificity, and 82.00% F1-score. The accuracy and efficiency of interpreting MRI images of the cervical spine may be enhanced by this approach. In this work [17] proposed to facilitate rapid automated and also objective diagnosis the evaluated models, VGG16 yielded the highest classification accuracy for LSS at 87.70%. There is a noticeable gap in reference to recent benchmarking studies involving ConvNeXtXLarge in spinal imaging classify and the increasing role of transformer architectures in medical image analysis. These emerging models have explained important promises in MRI classification often outperforming traditional CNN-based architectures in various domains, including spinal imaging. Integrating such models could further enhance the diagnostic performance of spinal imaging systems. Table 1 we obtained these summaries of previous research. Moreover, most current research does not involve datasets devoted to BMCLVB to publicly available, which limits reproducibility and further development. Compared to these earlier efforts, our study contributes a novel, dedicated MRI dataset of BMCLVB cases, leverages ConvNeXt XLarge used feature extraction and systematically evaluates multiple ML models classification. This methodology helps solve the problem of insufficient studies that directly focus on BMCLVB and proves the possibility of applying such practices to clinical decision support systems.

Table 1. Overview of existing ML and DL approaches for spinal image analysis

No.	Year	Dataset	Technique	Application	Performance (%)	Weakness
[12]	2024	MAGNETOM Skyra from siemens (Germany) and discovery 750 w from GE (United States)	Fine-tuned MedSAM	MRI image segmentation for spinal region analysis	AUC=95.00; accuracy=89.51; precision=87.07; recall=98.83; and F1-score=93.00	Focused only on segmentation; not evaluated for classification; no BMCLVB-specific data
[13]	2024	Achieva; Philips Healthcare, Best, the Netherlands	DL with attention UNet	Spinal cord lesion segmentation in MRI	AUC=87.00; sensitivity=80.00; and specificity=88.00.	No classification task; no transformer-based model comparison
[14]	2025	Kaggle Biomechanical features of orthopedic patients	DT, SVM, and NN	Classification of normal vs. abnormal spine biomechanics	Accuracy=90.83 and precision=91.86	Does not address marrow lesions; dataset not specific to lumbar vertebrae
[15]	2022	Industrial Technology Research Institute (ITRI), in collaboration with Taipei Medical University Hospital (TMUH)	LumbarNet U-Net	Lumbar spine segmentation in MRI for clinical diagnosis	Accuracy=88.83	Limited to segmentation; no radiomics or DL classifier benchmarking
[16]	2021	AO Spine Cervical Spondylotic Myelopathy North America (CSM-NA) trial (ClinicalTrials.gov: NCT00285337)	CNNs	Diagnosis of cervical spondylotic myelopathy	AUC=94.00; sensitivity=88.00; specificity=89.00; and F1-score=82.00	Targets cervical spine only; no feature extraction or ML comparison
[17]	2023	Lumbar spinal stenosis (LSS)	LSS-VGG16	Detection and classification of lumbar spinal stenosis	Accuracy=87.70	No deep feature extraction; lacks benchmarking with newer architectures

3. PROPOSED MODEL

The central goal of this research is to create a novel MRI dataset (BMCLVB) that facilitates the design of hybrid ML models. The experimentation includes successive stages of model training and validation as shown in Figure 1.

3.1. Collected dataset (BMCLVB)

The 1,018 MRI images from 134 patients who were assessed for bone marrow changes at Al-Kafeel Super Specialty Hospital (KSSH) in Iraq between April 10, 2022, and September 2, 2023, make up the BMCLVB dataset. T1-weighted, T2-weighted, and fat-suppressed sequences were all part of the MRI procedure and they were all obtained at a resolution of 256×256 pixels and a slice thickness of 5 mm. Table 2 summarizes the distribution of images by class.

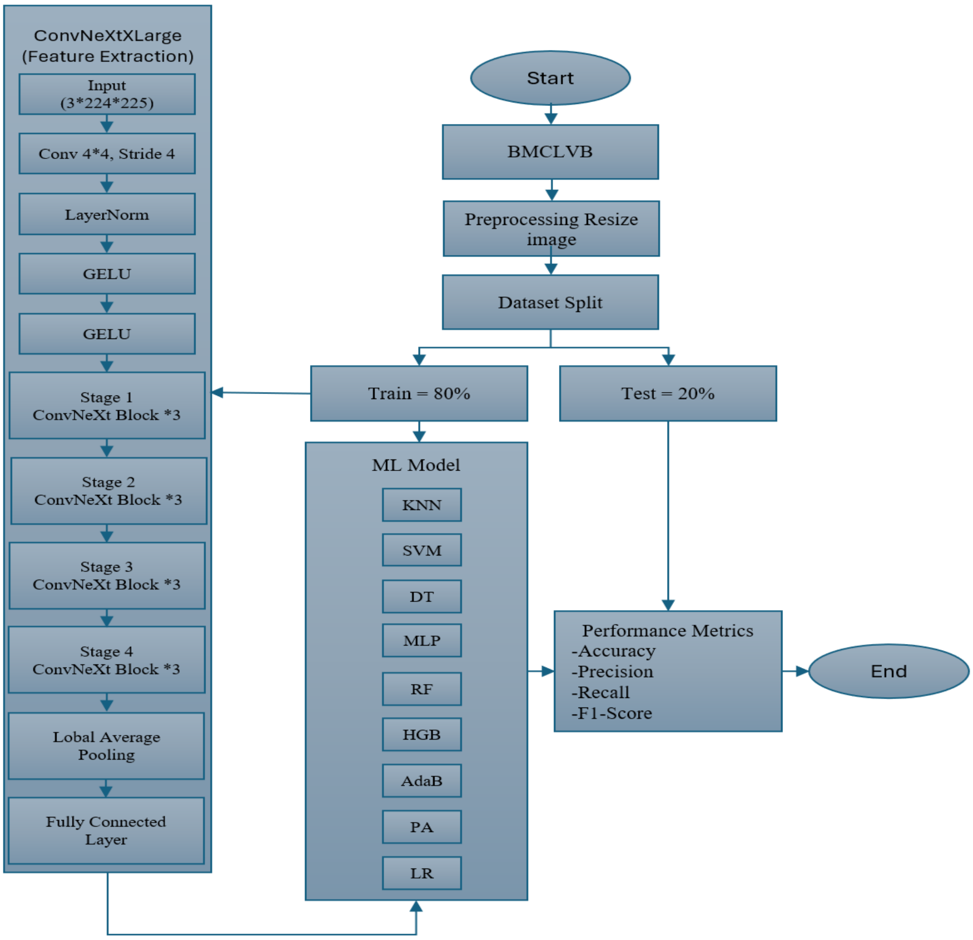


Figure 1. The flowchart of the proposed approach

Table 2. Class distribution of the BMCLVB MRI dataset (normal vs. abnormal)

Class	Number of images
Abnormal	560
Normal	458
Total	1018

The images were saved in DICOM format and anonymised for processing. The patients' ages ranged from 18 to 80 years. The study followed institutional review board (IRB) procedures and ethical standards outlined in the Declaration of Helsinki for studies involving participants. According to the study's findings, BMCLVB is diagnosed using clinical signs and symptoms, blood testing, MRI scans bone inspections. The conventional therapy includes long-term antibiotic or antifungal medication and, in certain circumstances, surgery to empty the abscess or remove the injured bone. Figure 2 presents examples of abnormal and normal MRI images of BMCLVB included in the dataset.

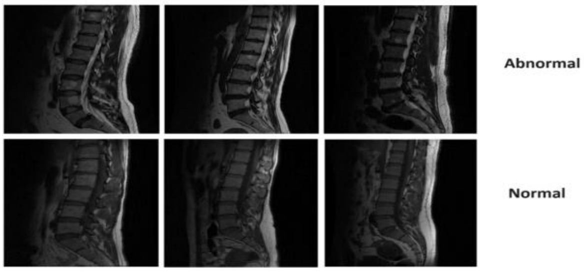


Figure 2. BMCLVB, abnormal, and normal MRI image

3.2. Pre-processing and resize images

The radiological images export were about 527 KB in size and formatted in DICOM with image dimensions of 512×512 pixels. To improve patient confidentiality and compatibility with most programming standard libraries, the images were converted to JPG format, resulting in a file size of (92) KB. The dataset is normalized across resizes and attributes, and all photos are scaled to 224×224. In this preprocess step ensures this all image inputs maintain uniform size and characteristics thereby aid consistent feature extraction in subsequent analysis.

3.3. Dataset split

The training and testing portions of the dataset were separated. The model was developed using the training data, and an invisible testing set was employed to evaluate its accuracy, followed by performance assessment on an independent test dataset. We split the dataset (1,018 instances) into two subsets: 80% (cases) for training and 20% (204), which were not subjected to training or testing and were used to evaluate the models' overall performance.

3.4. Feature extraction using ConvNeXt XLarge

The ConvNeXt XLarge network architecture used in this study is multiple convolutional layers make for a deep CNN, enabling powerful analysis of imaging data. In a given input images of three channels to form a 224×224×3 input, this network can undergo feature extraction as well as hierarchal representation learning. The last steps reveal another important stage that includes normalization of the layers, which is mandatory for stabilizing the training process and the convergence phase. Here, the output dimensions are (7,7, 4096), where it looks like the spatial dimensions are a lot smaller while the feature depth is quite large. This is possible because the 4,096 channels capture the various features within the images and the patterns and semantic information related to them are fully captured. The final layers enable the model to learn discriminative patterns, making ConvNeXt XLarge highly suitable for complex image recognition tasks, including medical imaging applications like MRI classification.

3.5. Machine learning model

In this section, nine several ML techniques were used classification spinal status (normal or abnormal) based extracted features from the dataset. The chosen algorithms include a mix of linear and non-linear classifiers, tree-based methods, ensemble techniques, and neural network-based models. These algorithms were select on provided comprehensive performance comparison across different learning. Given the limited number of instances dataset (1,018 in total) an 80:20 split was applied to create the training and testing sets, respectively.

3.5.1. K-nearest neighbors

KNN is a popular and successful classification method that divides data according to a specified distance metric in the feature space. The classification is based on how close a sample is to its KNN. A majority vote among these neighbors determines the final class assignment [18].

3.5.2. Support vector machine

The SVM is a powerful supervised learning method commonly applied to both classification and regression problems. SVM in classification allocates unknown information into one of the given classes by measuring data points through the vector of n-space dimensions having the number of attributes equal to ‘n’. The algorithm then determines the hyperplane that determines the maximum margin separating the two classes or more [19].

3.5.3. Decision tree

DT employ supervised learning algorithms which require the target variable to be predefined. They are particularly effective for exploratory tasks such as classification as they can model a wide range of functional forms and can approximate any functional form, thus supporting both continuous and categorical input/output types. The fundamental principle of the algorithm is to recursively partition the dataset into two or more homogeneous groups, based on the feature or input variable that offers the greatest discriminatory power [20].

3.5.4. Multilayer perceptron

MLP contains multiple perceptron because it is composed of more than one. The true computational power of MLP exists between its input and output layers in various hidden layers, which, along with one hidden layer enable any continuous function approximation [21].

3.5.5. Random forest
RF is an ensemble learning technique widely used for classification, regression, and other predictive tasks. It functions via generating numerous DT and aggregating their results taking the majority vote for classification and the mean value for regression [22].

3.5.6. Histogram-based gradient boosting
HGB is an advanced gradient boosting approach that supports classification and regression. It incrementally improves models to optimize differentiable loss functions and applies histogram binning to speed up computation [23].

3.5.7. Adaptive boosting
AdaBoost is a ML meta-algorithm proposed by Yoav Freund and Robert Schapire. The method works by sequentially training multiple weak learners each focusing on the instances misclassified via their predecessors and the output of these weak learner is then combined of a single boosted classifier through a weighted voting mechanism where weights are assigned based in each learner’s accuracy [24].

3.5.8. Passive aggressive
PA techniques are online learning techniques that remain passive for proper classification results but become aggressive in the case of a mistake, updating and modifying. It's appropriate for large-scale learning and may be utilized for binary or multiclass categorization [25].

3.5.9. Logistic regression
LR is a method of statistics for determining the validity of a given dataset in which one or more explanatory factors influence the variable of interest. A metric is an ordered quantitative variable that indicates the consequence of meeting a dual variable with just two alternatives. It has been employed in a variety of sectors, including social sciences, marketing, and medicine, among others [26].

3.6. Performance metrics
The experiments were performed on a system with two different Ryzen 7 (5800H 3.20) GHz CPUs and GPUs (T4 x2 and 16 GB of RAM). According to the above flow diagram, every step took place on Kaggle. Evaluation was conducted using metrics outlined in (1)-(4), the classification results were assessed by comparing the actual results with the model outcomes. These metrics are formulated as follows, and their calculation is given below, the greater the value, the better model's performance. True positive (TP) stands for correctly classified normal samples, true negative (TN) for correctly classified BMCLVB samples, false positive (FP) for BMCLVB samples categorized under the normal class, and false negative (FN) for normal samples misclassified as BMCLVB. Normal samples are considered positive for this study, whereas BMCLVB samples are considered negative.

$$Accracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

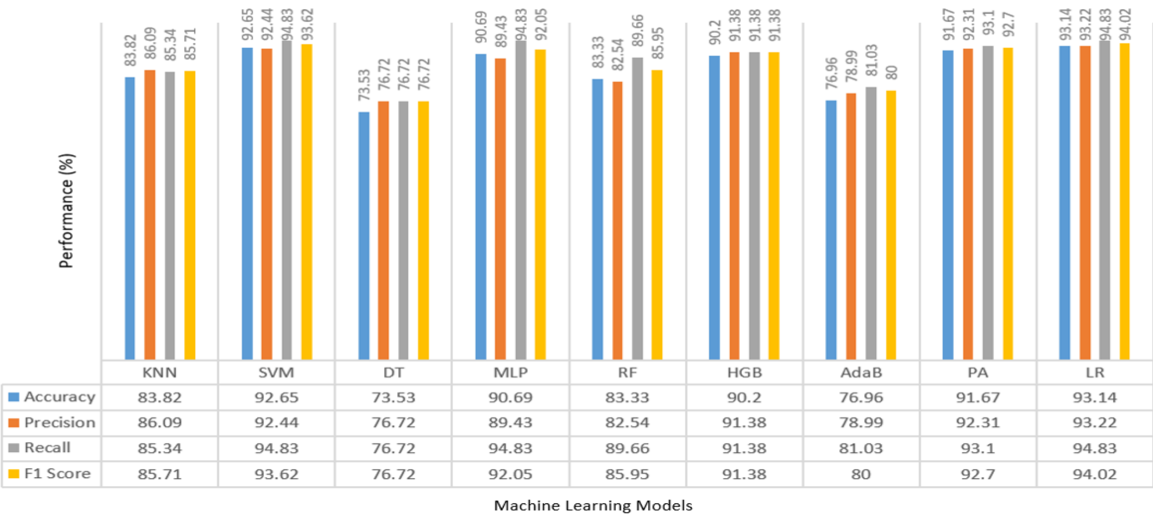
$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FP} \tag{3}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

4. RESULTS AND DISCUSSION
CNN models were trained using BMCLVB MRI scans. Following this, the models underwent training and testing based on the extracted features. The CNN architecture employed was ConvNeXt-XLarge, which generated 4096-dimensional feature vectors derived from layer normalization (the penultimate layer of the network). These feature vectors were subsequently used input for several ML classifiers, including KNN, SVM, DT, MLP, RF, HGB, AdaBoost, PA, and LR. As shown in Figure 3, LR achieved the highest performance on all evaluation metrics accuracy 93.14%, precision 93.22%, recall 94.83%, and an F1-score 94.02%. The high accuracy indicates good class discrimination while the strong precision and recall suggest minimal false positives and an excellent ability to identify true positives and also the balanced F1-score further confirms LR reliability in maintaining precision and recall balance. SVM also exhibited strong

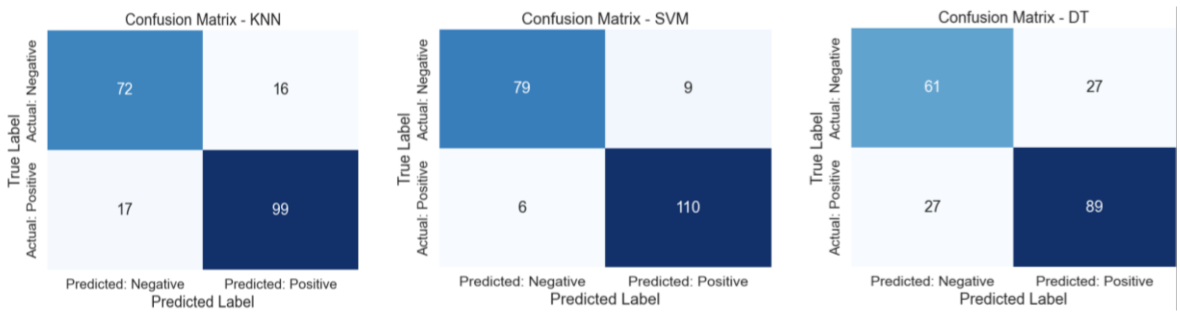
performance across all metrics of accuracy of 92.65%, precision of 92.44%, recall of 94.83%, F1-score of 93.62%, making it a highly dependable classifier for BMCLVB categorization and similarly, the MLP model demonstrated high accuracy of 90.69%, and F1-score of 92.05, though and recall was marginally lower than SVM and LR. RF did about as well as KNN but not quite as well as SVM, LR, MLP, and balanced performance implies that it could be useful in situations where interpretability and ensemble learning are important. Conversely the DT exhibited the weakest performance accuracy of 73.53% and F1-score of 76.72%, struggling with dataset complexity and capturing only simplistic patterns. AdaBoost and PA classifiers delivered moderate results with PA achieving accuracy of 91.67% and F1-score of 92.7%. HGB come out as a top contender matching LR, SVM, and MLP in performance of accuracy of 90.2%, F1-score of 91.38% and the lastly KNN performed reasonably well accuracy of 83.82%, F1-score of 85.71% but remained inferior compared to the best-performing models.



	KNN	SVM	DT	MLP	RF	HGB	AdaB	PA	LR
Accuracy	83.82	92.65	73.53	90.69	83.33	90.2	76.96	91.67	93.14
Precision	86.09	92.44	76.72	89.43	82.54	91.38	78.99	92.31	93.22
Recall	85.34	94.83	76.72	94.83	89.66	91.38	81.03	93.1	94.83
F1 Score	85.71	93.62	76.72	92.05	85.95	91.38	80	92.7	94.02

Figure 3. Performance evaluation of all models

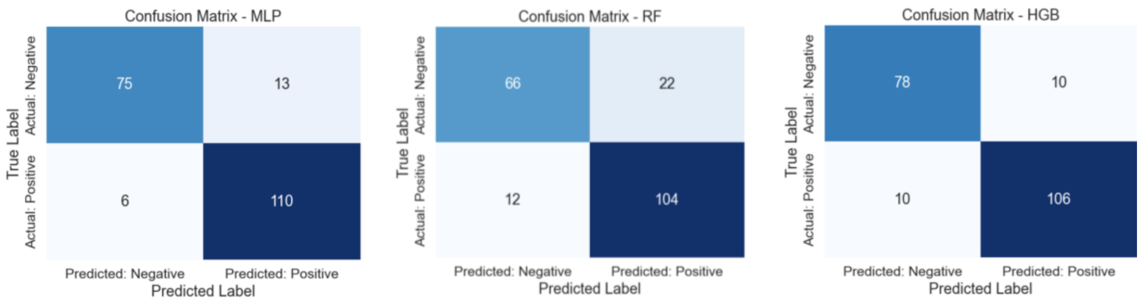
The confusion matrices for KNN, SVM, and DT classifiers on the BMCLVB test dataset. The matrices show TP, TN, FP, and FN results for each model and LR in particular, excelled in TP and TN, indicating reliable classification of BMCLVB cases. In contrast, the DT model has a higher number of FP and FN and therefore, they are more likely to misclassify. Figures 4(a)–(c) give confusion matrices of three ML classifiers across the BMCLVB test dataset. That is, Figure 4(a) illustrates KNN's classifier having moderate classification accuracy but with observed FN. Figure 4(b) presents the confusion matrix for SVM's classifier with good discrimination between normal and abnormal cases having very minimal misclassifications. Figure 4(c) shows results from the DT classifier having the worst performance with high false-positive and false-negative rates across all models.



(a) (b) (c)

Figure 4. Confusion matrices of ML methods; (a) KNN, (b) SVM, and (c) DT

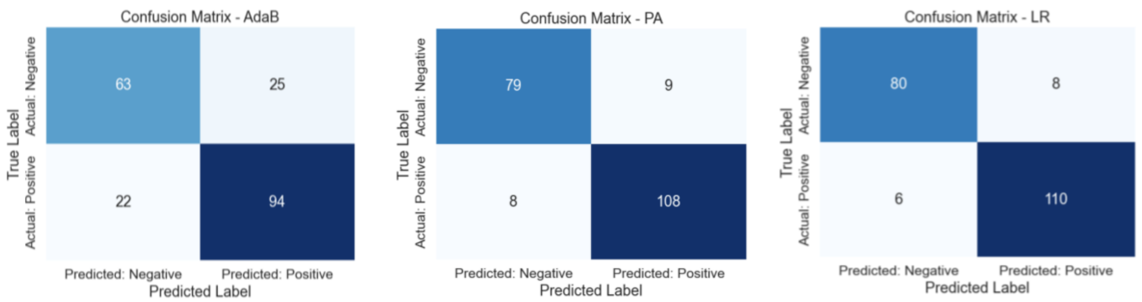
Figures 5(a)–(c) show confusion matrices for three ensemble and neural network–based models. Whereas Figure 5(a) depicts the MLP classifier having good abnormal case recall though slightly lower precision, Figure 5(b) depicts RF outcomes having balanced results and moderate misclassification. Figure 5(c) depicts HGB classifier having strong classification with minimum error through both classes.



(a) (b) (c)

Figure 5. Confusion matrices of ML methods; (a) MLP, (b) RF, and (c) HGB

The confusion matrices of other classifiers are provided in Figures 6(a)–(c). Figure 6(a) shows AdaBoost's outcomes with moderate ability and slightly high misclassification. Figure 6(b) outlines the PA classifier with good precision and strong recall. Figure 6(c) contains the LR model having best general performance with minimum false classification and most balanced results.



(a) (b) (c)

Figure 6. Confusion matrices of ML methods; (a) AdaBoost, (b) PA, and (c) LR

This will essentially imply that the success rate in classification of the models will also vary from one class to another becouse to the variation in the numbers of images of each class. Therefore, classification accuracy for each class also calculated and improved classification performance as demonstrated by ConvNeXt XLarge+LR, directly supports clinical decision-making by enhancing diagnostic accuracy. This can reduce misdiagnosis in spinal imaging, especially in overburdened radiology departments. Automated detection tools can also assist in early intervention, especially for patients with comorbidities such as osteoporosis or cancer, thus streamlining care pathways. LR and SVM models outperformed others likely due to their robustness in high-dimensional, structured feature spaces, which align well with ConvNeXt XLarge dense features. Conversely, DT underperformed due to its sensitivity to noise and lack of regularization, which limits its ability to generalize in complex medical imaging data. Table 3 illustrates the general classification performance of the models studied and the classification success of each class, respectively.

This study is subject to certain limitations: i) the dataset size, although novel and sizable in BMCLVB context, remains limited for broader generalization; ii) the lack of external validation may introduce dataset-specific bias, and iii) absence of data augmentation might have restricted model generalizability. Future work will address these issues by incorporating external datasets, more sophisticated augmentation, and transfer learning.

Table 3. Performance comparison of ML BMCLVB classes

Method	Class	Precision	Recall	F1-score	Support	Average accuracy
KNN	Normal	0.81	0.82	0.81	88	
	Abnormal	0.86	0.85	0.86	116	0.84
SVM	Normal	0.93	0.90	0.91	88	
	Abnormal	0.92	0.95	0.94	116	0.92
DT	Normal	0.69	0.69	0.69	88	
	Abnormal	0.77	0.77	0.77	116	0.73
MLP	Normal	0.93	0.85	0.89	88	
	Abnormal	0.89	0.95	0.92	116	0.91
RF	Normal	0.85	0.75	0.80	88	
	Abnormal	0.83	0.90	0.86	116	0.84
HGB	Normal	0.89	0.89	0.89	88	
	Abnormal	0.91	0.91	0.91	116	0.90
AdaBoost	Normal	0.74	0.72	0.73	88	
	Abnormal	0.79	0.81	0.80	116	0.77
PA	Normal	0.91	0.90	0.90	88	
	Abnormal	0.92	0.93	0.93	116	0.92
LR	Normal	0.93	0.91	0.92	88	
	Abnormal	0.93	0.95	0.94	116	0.94

Another startling problem with the same is that MRIs often identify 75–80% of spinal cord damage. Research in the Journal of Neurotrauma highlights the essential need for MRI in evaluating acute spinal cord injury (SCI). This emphasizes how important MRI is for a prompt and precise diagnosis of SCI [27], [28]. The length of diagnosis is another significant component of the study that adds to the body of knowledge. Our study accomplishes its objective by cutting down on the amount of time radiology professionals require to diagnose patients. Our research demonstrates that the proposed approach effectively reduces the time required for radiology professionals to diagnose patients. Throughout this research, we have emphasized that our primary goal is to develop systems that support clinicians in improving diagnostic accuracy and efficiency. Experimental results demonstrate that the intensity of persistent low back pain is strongly correlated with with BMCLVB, articularly in patients with comorbidities such as cancer, and osteoporosis, and in high-demand clinical settings, radiologists may miss subtle indicators due to workload pressures and fatigue. As summarized in Table 4, our results highlight strong performance of the propos model on addressing these challenges.

Table 4. Comparison of detection performance with previous ML and DL studies

No.	Year	Technique	Performance (%)
[12]	2024	Fine-tuned MedSAM	AUC=95.00; accuracy=89.51; precision=87.07; recall=98.83; and F1-score=93.00.
[13]	2024	DL with attention UNet	AUC=87.00; sensitivity=80.00; and specificity=88.00.
[14]	2025	DT, SVM, and neural networks	Accuracy=90.83 and precision=91.86.
[15]	2022	LumbarNet U-Net	Accuracy=88.83.
[16]	2021	CNNs	AUC=94.00; sensitivity=88.00; specificity=89.00; and F1-score=82.00.
[17]	2023	LSS-VGG16	Accuracy=87.70.
Our study	2025	ConvNeXtXLarge-LR	Accuracy=93.14; precision=93.22; recall=94.83; and F1-score=94.02.

5. CONCLUSION

In conclusion, this study demonstrates the potential of MRI-based classify of BMCLVB and explains that such changes can be effectively identified by application of ML techniques. By combining ConvNeXt XLarge for feature extraction with LR classifiers our approach achieved best performance across all evaluation metrics. The results highlight the potential of this hybrid methodology to improve diagnostic accuracy and accelerate clinical decision-making to support early detection of BMCLVB related conditions. Accurate classification of bone marrow change is critical for diagnosing a range of diseases, including degenerative disc disease, infections, and neoplasms. The proposed model provides a dependable and cost effective that can assist clinicians in treatment planning and monitoring disease progression. The powerful performance of our system shown and feasibility for merging into clinical workflows offering valuable findings derived from MRI. Automated classification tools like our approach have the potential to enhance efficiency reduce diagnostic minimize and improve patient results. our findings highlight the potential of ML-based techniques to medical imaging interpretation with implications for better diagnosis, planning of therapy, and patient results in lumbar vertebrae pathology. Further study and collaboration are required to realize the full potential of these strategies in clinical practice. Many possible avenues for further study were offered including tests utilizing different DL techniques.

ACKNOWLEDGMENTS

We acknowledge the Al-Kafeel Super Specialty Hospital (KSSH), Dr. Ali Kanj, Dr. Hisham Hassan Abd, Dr. Mohamed Abdelreda, and the Electronics Department's Mr. Kamal Al-Deen Mahdi Hussein and Mr. Mohammed Abbood Kadhim for their assistance in gathering data. The trained code models are available on GitHub at: yasserhessein/Bone-Marrow-Changes-in-Lumbar-Vertebrae-with-ConvNeXtXLarge-Machine-Learning.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Yasir Hussein Shakir	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			✓
Tiong Sieh Kiong	✓	✓		✓	✓	✓	✓			✓		✓	✓	
Chai Phing Chen	✓	✓		✓	✓	✓	✓			✓		✓	✓	
Sachin Sharma Ashok Kumar	✓	✓		✓	✓	✓	✓			✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state that there are no conflicts of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] R. Chou *et al.*, "Diagnosis and Treatment of Low Back Pain: A Joint Clinical Practice Guideline from the American College of Physicians and the American Pain Society," *Annals of Internal Medicine*, vol. 147, no. 7, pp. 478–491, Oct. 2007, doi: 10.7326/0003-4819-147-7-200710020-00006.

[2] N. M. Moll *et al.*, "Multiple Sclerosis Normal-Appearing White Matter: Pathology–Imaging Correlations," *Annals of Neurology*, vol. 70, no. 5, pp. 764–773, Nov. 2011, doi: 10.1002/ana.22521.

[3] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019, doi: 10.1186/s40537-019-0197-0.

[4] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[5] S. Kumar and M. Singh, "Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools," *Big Data Mining and Analytics*, vol. 2, no. 1, pp. 48–57, 2019, doi: 10.26599/BDMA.2018.9020031.

[6] L. M. Ang, K. P. Seng, G. K. Ijemaru, and A. M. Zungeru, "Deployment of Internet of Vehicles for Smart Cities: Applications, Architecture, and Challenges," *IEEE Access*, vol. 7, pp. 6473–6492, 2019, doi: 10.1109/ACCESS.2018.2887076.

[7] B. P. L. Lau *et al.*, "A Survey of Data Fusion in Smart City Applications," *Information Fusion*, vol. 52, pp. 357–374, 2019, doi: 10.1016/j.inffus.2019.05.004.

[8] Y. Wu *et al.*, "Large Scale Incremental Learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, doi: 10.1109/CVPR.2019.00046.

[9] A. Mosavi, S. Shamshirband, E. Salwana, K. W. Chau, and J. H. M. Tah, "Prediction of Multi-Inputs Bubble Column Reactor Using a Novel Hybrid Model of Computational Fluid Dynamics and Machine Learning," *Engineering Applications of Computational Fluid Mechanics*, vol. 13, no. 1, pp. 482–492, 2019, doi: 10.1080/19942060.2019.1613448.

[10] H.-N. Zhu, Y.-F. Guo, Y. Lin, Z.-C. Sun, X. Zhu, and Y. Li, "Radiomics Analysis of Thoracic Vertebral Bone Marrow Microenvironment Changes Before Bone Metastasis of Breast Cancer Based on Chest CT," *Journal of Bone Oncology*, vol. 50, 2025, doi: 10.1016/j.jbo.2024.100653.

[11] D. S. W. Lim *et al.*, "Improved Productivity Using Deep Learning–Assisted Reporting for Lumbar Spine MRI," *Radiology*, vol. 305, no. 1, pp. 160–166, 2022, doi: 10.1148/radiol.220076.
[12] J. Xie *et al.*, "MRI Radiomics-Based Decision Support Tool for a Personalized Classification of Cervical Disc Degeneration: A Two-Center Study," *Frontiers in Physiology*, vol. 14, 2023, doi: 10.3389/fphys.2023.1281506.
[13] Y. Lin, S. C. W. Chan, H. Y. Chung, K. H. Lee, and P. Cao, "Deep Neural Network for MRI Spinal Inflammation in Axial Spondyloarthritis," *European Spine Journal*, vol. 33, no. 11, pp. 4125–4134, 2024, doi: 10.1007/s00586-023-08099-0.
[14] D. Nasef, V. Sawiris, P. Girgis, and M. Toma, "Machine-Learning-Based Biomechanical Feature Analysis for Orthopedic Patient Classification with Disc Hernia and Spondylolisthesis," *BioMedInformatics*, vol. 5, no. 1, 2025, doi: 10.3390/biomedinformatics5010003.
[15] G. M. Trinh *et al.*, "Detection of Lumbar Spondylolisthesis from X-ray Images Using Deep Learning Network," *Journal of Clinical Medicine*, vol. 11, no. 18, 2022, doi: 10.3390/jcm11185450.
[16] Z. Merali, J. Z. Wang, J. H. Badhiwala, C. D. Witiw, J. R. Wilson, and M. G. Fehlings, "A Deep Learning Model for Detection of Cervical Spinal Cord Compression in MRI Scans," *Scientific Reports*, vol. 11, no. 1, May 2021, doi: 10.1038/s41598-021-89848-3.
[17] S. Altun, A. Alkan, and İ. Altun, "LSS-VGG16: Diagnosis of Lumbar Spinal Stenosis with Deep Learning," *Clinical Spine Surgery*, vol. 36, no. 5, pp. E180–E190, 2023, doi: 10.1097/BSD.0000000000001418.
[18] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN Model-Based Approach in Classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, Catania, Italy, 2003, pp. 986-996, doi: 10.1007/978-3-540-39964-3_62.
[19] D. Keerthana, V. Venugopal, M. K. Nath, and M. Mishra, "Hybrid Convolutional Neural Networks with SVM Classifier for Classification of Skin Cancer," *Biomedical Engineering Advances*, vol. 5, pp. 1-8, Jun. 2023, doi: 10.1016/j.bea.2022.100069.
[20] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of Use of Decision Tree Algorithms in Machine Learning," in *Proceedings of the IEEE Control and System Graduate Research Colloquium (ICSGRC 2011)*, 2011, pp. 986-996, doi: 10.1109/ICSGRC.2011.5991826.
[21] M. Desai and M. Shah, "An Anatomization on Breast Cancer Detection and Diagnosis Employing Multi-Layer Perceptron Neural Network (MLP) and Convolutional Neural Network (CNN)," *Clinical eHealth*, vol. 4, pp. 1–11, 2021, doi: 10.1016/j.ceh.2020.11.002.
[22] C. Zhang, Y. Li, Z. Yu, and F. Tian, "Feature Selection of Power System Transient Stability Assessment Based on Random Forest and Recursive Feature Elimination," in *Proceedings of the Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2016, pp. 1264-1268, doi: 10.1109/APPEEC.2016.7779696.
[23] A. Guryanov, "Histogram-Based Algorithm for Building Gradient Boosting Ensembles of Piecewise Linear Decision Trees," in *Analysis of Images, Social Networks and Texts: 8th International Conference (AIST 2019)*, Kazan, Russia, Jul. 2019, pp. 36–47, doi: 10.1007/978-3-030-37334-4_4.
[24] P. Favaro and A. Vedaldi, "AdaBoost," in *Computer Vision: A Reference Guide*, Springer International Publishing, 2021, pp. 36–40, do: 10.1007/978-3-030-63416-2_663.
[25] L. Gao, W. Zhang, and Q. Tang, "Passive Aggressive Algorithm for Online Portfolio Selection with Piecewise Loss Function," in *Proceedings of the 9th International Conference on Advanced Data Mining and Applications (ADMA 2013)*, Hangzhou, China, 2013, pp. 360–371, doi: 10.1007/978-3-642-53917-6_32.
[26] M. Maalouf, "Logistic Regression in Data Analysis: An Overview," *International Journal of Data Analysis Techniques and Strategies*, vol. 3, no. 3, pp. 281–299, 2011, doi: 10.1504/IJDATS.2011.041335.
[27] H. Yamamoto, H. Nakagawa, T. Yamada, K. Iwata, T. Okumura, and D. Hoshino, "Magnetic Resonance Imaging of Acute Spinal-Cord Injury," *CT Kenkyu*, vol. 14, no. 2, pp. 155–160, 1992.
[28] C. O. Walsh, S. I. Ziniel, H. K. Delichatsios, and D. S. Ludwig, "Nutrition Attitudes and Knowledge in Medical Students After Completion of an Integrated Nutrition Curriculum Compared to a Dedicated Nutrition Curriculum: A Quasi-Experimental Study," *BMC Medical Education*, vol. 11, no. 1, p. 582011, doi: 10.1186/1472-6920-11-58.

BIOGRAPHIES OF AUTHORS

Yasir Hussein Shakir is a software engineer who obtained his B.Sc. degree in software engineering from Baghdad College of Economic Sciences University in 2014. He further pursued his education and received his M.Sc. degree in Computer and Communication, specializing in Computer Programming, from the Faculty of Engineering at the Islamic University of Lebanon (IUL) in 2018. His areas of interest include data mining, medical image processing, medical electronic systems, machine learning, deep learning, and artificial intelligence. Currently, he is a Ph.D. student in the Department of Engineering at Universiti Tenaga Nasional (UNITEN) in Malaysia. He can be contacted at email: yasserhessein19855@gmail.com.

Tiong Sieh Kiong is a Senior Professor and a Senior Member of the IEEE, currently serving in the College of Engineering at Universiti Tenaga Nasional (UNITEN). He also holds the position of Director at the Institute of Sustainable Energy (ISE), UNITEN. He received his B.Eng. (Hons), M.Sc., and Ph.D. in Electrical and Electronic Engineering. His research interests include renewable energy, artificial intelligence, data analytics, and communication systems. He is a Professional Engineer registered with the Board of Engineers Malaysia (BEM) and a Member of the Institute of Electrical and Electronics Engineers (IEEE). He can be contacted at email: siehkiong@uniten.edu.my.

Chai Phing Chen is a senior lecturer in Universiti Tenaga Nasional (UNITEN), Malaysia. She has been associated with technical education for more than ten years and actively participated in various UNITENUNITEN projects which involve machine learning. Her project field contains predicting for gas emission from power plant; prediction for short-term wind speed; health condition analysis for HV circuit breaker; fault detection for switchgear; fault detection for transformers; heat waste recovery system via thermoelectric generator; and nontechnical losses detection. She can be contacted at email: chencp@uniten.edu.my.

Sachin Sharma Ashok Kumar is currently a senior lecturer in the College of Engineering, Taylor's University, Malaysia and also an associate professor at Lincoln University College (LUC), Faculty of Engineering, Malaysia. He received both of his B.Sc. degree (Hons.) and M.Sc. (Hons.) in Mechanical Engineering minor in Materials Science from Wichita State University, USA in the years 2011 and 2012, respectively. He received his Ph.D. in Advanced Materials Science Engineering at the University of Malaya in 2023. In the past decade, his research has focused on the synthesis of graphene/graphene oxide films incorporated with reinforced composites, supercapattery, batteries, solar cells, fuel cells, hydrogen storage, polymer nanocomposites, corrosion coatings, and 3D composites. for numerous engineering applications. His current research involves the synthesis of super-hydrophobic graphene-based polymer nanocomposite coatings to enhance corrosion resistance. He has published over 40 articles in high ranked journals, participated in international conferences/exhibitions as a keynote speaker and has received several awards at both local and international levels. He is currently a member registered with the Board of Engineers Malaysia (BEM). He can be contacted at email: sachin.sharma@taylors.edu.my.