IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 5, 2022, accepted April 27, 2022, date of publication May 9, 2022, date of current version May 23, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3173424

HipXNet: Deep Learning Approaches to Detect Aseptic Loos-Ening of Hip Implants Using X-Ray Images

TAWSIFUR RAHMAN[1], AMITH KHANDAKAR[1], (Senior Member, IEEE), KHANDAKER REAJUL ISLAM[2], MD MOHIUDDIN SOLIMAN[3], MOHAMMAD TARIQUL ISLAM[3], (Senior Member, IEEE), AHMED ELSAYED[4,5], YAZAN QIBLAWEY[1], (Member, IEEE), SAKIB MAHMUD[1], ASHIQUR RAHMAN[6], FARAYI MUSHARAVATI[7], ERFAN ZALNEZHAD[8], AND MUHAMMAD E. H. CHOWDHURY[1], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Qatar University, Doha, Qatar
[2]Department of Orthodontics, Bangabandhu Sheikh Mujib Medical University, Dhaka 1000, Bangladesh
[3]Department of Electrical, Electronics and Systems Engineering, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia
[4]Orthopaedic Department, Hamad Medical Corporation (HMC), Doha, Qatar
[5]Clinical Orthopaedic Surgery, Weill Cornell Medical College, Doha, Qatar
[6]Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Sendai 980-8577, Japan
[7]Department of Mechanical and Industrial Engineering, Qatar University, Doha, Qatar
[8]Department of Biomedical Engineering, The University of Texas at San Antonio, San Antonio, TX 78249, USA

Corresponding authors: Farayi Musharavati (farayi@qu.edu.qa) and Muhammad E. H. Chowdhury (mchowdhury@qu.edu.qa)

This work was supported in part by the Qatar National Research Fund (QNRF) under Grant NPRP11S-0102-180178, and in part by the Qatar National Library.

ABSTRACT Radiographic images are commonly used to detect aseptic loosening of the hip implant in patients with total hip replacement (THR) surgeries. These techniques of manual assessment by medical professionals can suffer from the drawback of low accuracy, poor inter-observer reliability, and delays due to the unavailability of experienced clinicians. Thus, the paper provides a reliable Deep Convolutional Neural Networks (DCNNs) based novel stacking approach (HipXNet) for detecting loosening of the hip implant using X-ray images. Two major investigations were done in this study. Firstly, the performance of four different state-of-the-art object detection YOLOv5 models was evaluated to detect the implant region from the hip X-ray images. Secondly, the study developed a stacking classifier using three different Convolutional neural networks (CNN) models to classify aseptic hip loosening and compared the performance with eight different state-of-the-art CNN networks. Moreover, one publicly accessible dataset with two sub-sets was created for these two experiments, where 200 hip implant X-ray images were collected and annotated by two expert radiologists for implant detection and 206 hip implant X-ray images were collected for loosening detection. YOLOv5m model outperformed the other variants of YOLOv5 to detect the implant region with the precision, recall, mean average precision $(mAP)_{0.5}$, $mAP_{0.5-0.95}$ of 100%, 100%, 100%, and 87.8%, respectively. Densenet201 CNN model outperformed other CNN models with the accuracy, precision, sensitivity, F1 score, and specificity of 94.66%, 94.66%, 94.66%, 94.66%, and 94.5%, respectively while the stacking technique with Random Forest meta learner classifier produced the best performance with the accuracy, precision, sensitivity, F1 score and specificity of 96.11%, 96.42%, 96.42%, 96.42%, and 96.74% respectively for loosening detection. The reliability of the performance was confirmed by the popular Score-CAM visualization. This study can help in the early and fast identification of hip implant loosening with the help of simple X-ray images and computed aided diagnosis.

INDEX TERMS Hip implant, aseptic loosening, total hip replacement, convolutional neural network, stacking technique.

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino.

I. INTRODUCTION

It has been reported that about 20% of people with an age higher than forty agonize because of bone degenerative diseases [1] like osteoporosis which leads to a worldwide request for procedures of total hip replacement (THR). Getting involved with osteoporosis for the aged population is inevitable. It is predicted that people aged 65 or higher are more susceptible and the rate of disease for them will be increased from 8.2% in 2018 to 17.6% in 2060 [2]. Functional failure of the implant may be followed by revision surgery, which is often painful and has a relatively low success rate [1], [3]. The lifetime of implants depends on (i) the type of materials used for implants, (ii) surgical techniques implemented, (iii) geometry of implant, (iv) patient's physical activity, and (v) patient's age. Of major concern to implant recipients is the life cycle of implants, which is currently limited to the order of 10-15 years. This relatively short life can be attributed to implant wear, loosening, and misalignment, which often cause pain and discomfort to the patient. Wear and corrosion due to the contact of the implants with other parts and body fluids generate debris. Soluble debris goes to blood and secretes through urine, however, particular debris gets accumulated in tissues, lymph, and bone marrow. This accumulated debris has short terms and long terms effects such as inflammation and cell tissue damage, hypersensitivity, chromosomal abrasion, and toxicity, in which both short and long-term effects result in revision surgery. Fibrous encapsulation due to the non-bonding of implants with surrounding tissues and inflammation (rejection) are the other reasons for implant failure. Osteoporosis, osteoarthritis, and trauma diseases are among the main reasons for replacements of the joints.

Despite the revolution in total hip arthroplasty (THA) in arthritis treatment, aseptic (mechanical) loosening always leads to joint failure and THR surgery [4]. Implant failure can be identified with radiolucent changes surrounding acetabular and femoral implants and progression of osteolysis [5]. In a previously asymptomatic total hip arthroplasty, new onset of pain could indicate implant loosening, infection, or both. Aseptic loosening-related pain is often increased by weight-bearing and range of motion, especially with internal and exterior rotations. When loosening happens early after surgery for no obvious reason, the infection should be investigated. Fever, chills, and restless pain are some of the symptoms that may accompany an infection. Unfortunately, aseptic loosening does not always cause discomfort to the patient. The loosening can be unpleasant with cemented acetabular components, which are only symptomatic in 10% of instances [6]. X-ray or computed tomography (CT) images of the hip area are commonly used for detecting the aseptic loosening of the hip implants by a medical expert as these can be done easily and readily. The foundation of aseptic loosening in radiological evaluation is the visual detection of radiolucent areas around the bone-cement or bone-prosthesis interface [7], [8]. Even in well-fixed implants immediately after surgery, thin sclerotic lines can be visible, which might indicate that the prosthesis is loose but in reality, it is not [8]. As a result, loosening is determined by the extent of radiolucent zones surrounding the implant and the change (progression) in appearance over time [9], [10]. This results in frequent hospital visits and prolonged patient follow-up to confirm the diagnosis of mechanical loosening while increasing patient morbidity and resource consumption. Aseptic loosening is typically indicated by loosening more than 2 mm or increased loosening on repeated radiographs [9], [10].

The human eye, especially on consecutive images, has a remarkable capacity for recognizing complicated patterns. The human aspect, on the other hand, has its own set of issues. Individuals notice patterns in different ways and give varying weights to different characteristics based on their unique experiences, making it challenging to translate what one practitioner sees visually into a set of ''rules'' that others can follow. As a result, there is significant scope for inter-and intra-subject variability and errors. Since the accurate quantification of aseptic loosening and correct identification of progression can reduce the consequence of the aseptic loosening, artificial intelligence (AI) based detection and quantification have a high potential to avoid subjective variability and error. Temmerman *et al.* [11] reported that X-ray images of the hip implant can be used to diagnose cementless femoral component loosening with the sensitivity and specificity of 50% and 89.5%, respectively. Inter-observer agreement was found to be very low (intraclass correlation coefficient (ICC) of -0.1). Chang *et al.* in [9] reported sensitivity and specificity of 83% and 82%, respectively. Temmerman *et al.* [12] showed that plain radiography has a sensitivity of 85% and a specificity of 78% for the diagnosis of cementless acetabular component loosening. There was a moderate inter-observer agreement, and an ICC of -0.53 was reported. Aseptic loosening is even more difficult to diagnose early. Khalily and Whiteside [13] showed that the presence of radiolucent lines around porous-coated femoral stems was 100 percent sensitive but only 55 percent specific for predicting the need for future revision (8-12 years post-surgery) at a 2-year follow-up. These variations in assessment have been questioned by Smith *et al.* [14]. Alternative imaging modalities including computed tomography, magnetic resonance imaging, ultrasound, bone scans, and arthrography can increase diagnostic accuracy, but they come with higher costs, some with ionizing radiation exposure, and the risk of contrast agents.

Recently, machine learning application for reliable automatic detection of abnormalities have become popular for COVID-19 detection using radiological images [15], [16], diabetic foot complication detection [17], tuberculosis detection [18] etc. In these applications, the machine has performed comparable to that of professional surgeons and radiologists, and even better than that of general practitioners [19]. To the best of the authors' knowledge, no previous work has reported the use of object detectors to detect the implant region from hip implant X-ray images so that the deep learning model can be trained to detect aseptic loosening accurately. This study first focuses on the detection of the implant

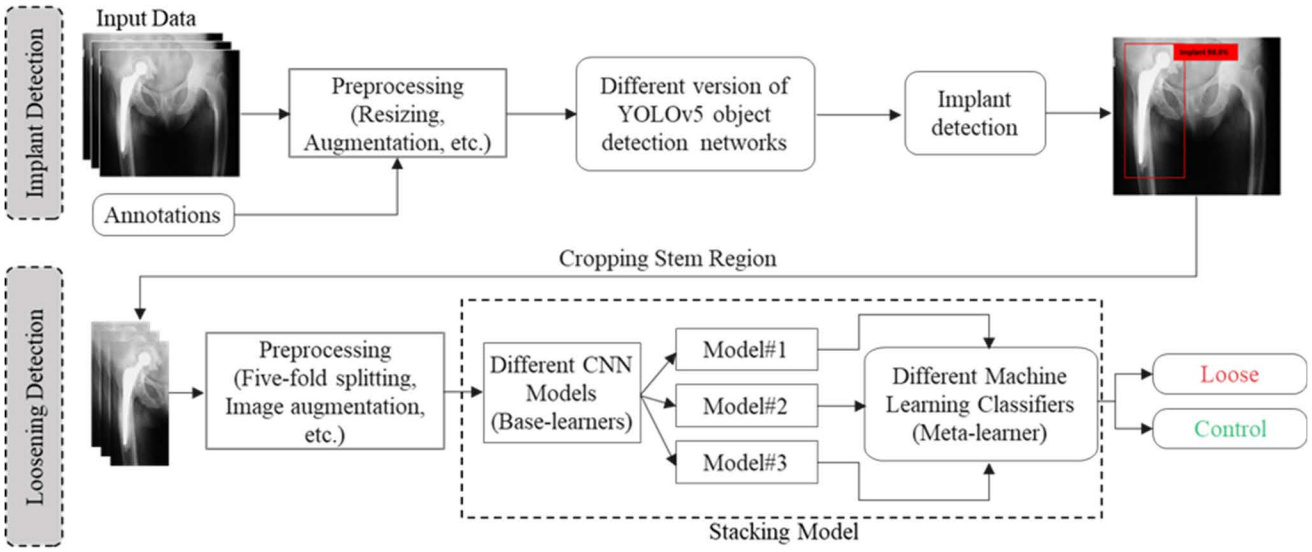

FIGURE 1. Methodology of the study.

area before loosening detection so that the implant region can be precisely used to identify whether the hip implant is loose or not. This study also reported saliency map visualization to confirm that the deep learning models are learning from the relevant region of interest for the classification. The major contributions of this study are:

- Firstly, we developed the first publicly accessible dataset of hip implant X-ray images with two sub-sets. One sub-set is made up of hip area X-ray images (single leg or both legs) with implant annotations and another sub-set is for hip loosening detection, where two classes (Control and Aseptic loosening) of X-ray images are available.
- We developed an object detection model based on YOLOv5 architecture for the detection of the implant region from the X-ray images and compared the performance with different versions of the YOLOv5 network.
- Then, we developed a stacking classifier using three different Convolutional neural networks (CNN) models to classify aseptic hip loosening and compared the performance with eight different state-of-the-art CNN networks.
- Finally, Score-CAM based visualization technique was used to display the saliency map of the most contributory area for loosening detection to confirm the reliability of the model.

The rest of the paper is divided into the following sections: Section 2 describes the dataset, pre-processing steps, and methodology of this study, while Section 3 provides the results and Section 4 discussed the results of two major experiments: implant localization and loosening detection from the hip implant X-ray images. Finally, section 5 concludes the study.

II. METHODOLOGY

The overall methodology of this study is illustrated using Figure 1. Two main experiments were carried out in this study. Firstly, four different versions of YOLOv5 models (such as YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x) were investigated to detect the implant regions from the hip implant X-ray images [20]. Secondly, the implant regions of the X-ray images were used for aseptic loosening detection using a novel stacking classifier and compared the performance of this classifier with different eight CNN networks. The stacking approach consists of two learners namely, base-learners and meta-learners. We used three different CNN models in base learners and the output of the base learners was predicted by a machine learning classifier as a meta-learner. Lastly, we evaluated the classification reliability using the Score-CAM technique. The experiments were done using PyTorch library with Python 3.7 on Intel® Xeon® CPU E5-2697 v4 @ 2,30GHz and 64 GB RAM, with a 16 GB NVIDIA GeForce GTX 1080 GPU. The section also provides the details of the dataset used, pre-processing steps applied, machine learning approaches adopted, performance metrics, and the visualization technique used in the study.

A. DATASET DESCRIPTION

Although there is a large number of THR surgery taking place all over the world, no publicly available hip implant X-ray image database for implant localization and aseptic loosening detection with control groups are present. This has motivated the authors to create such a dataset. A Kaggle dataset was created and made publicly available so that worldwide researchers can develop an AI-based model for computer-aided diagnosis and take benefit of this dataset. This dataset is made up of two sub-sets: one was for implant localization from the hip implant X-ray images, while the other one was for hip arthroplasty loosening detection. A detailed description of the Dataset, dataset preparation, and experiments are presented below.

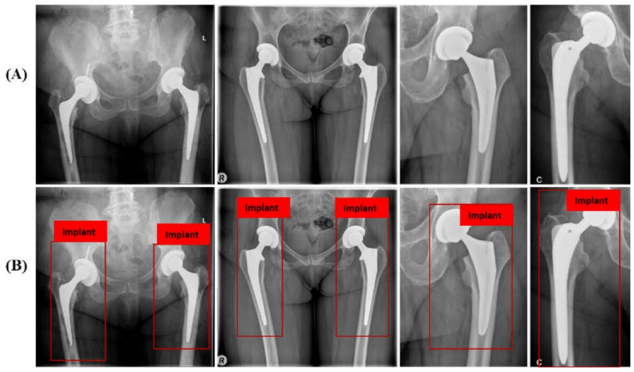

FIGURE 2. Sample images of implant detection dataset; (A) Hip implant X-ray images, and (B) Hip implant with Implant annotation.

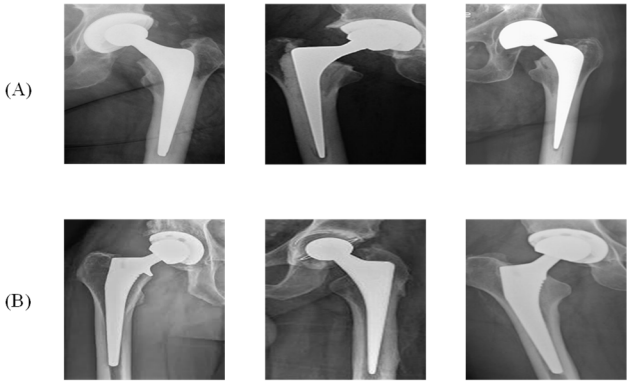

FIGURE 3. Sample hip implant X-ray images of loosening detection dataset: (A) Control, and (B) Aseptic Loosening.

1) IMPLANT DETECTION SUB-SET

Authors have collected and indexed this first sub-set of X-ray images of hip implants from different publicly available online medical sources such as medicine journals (articles) and radiology websites. All images should at least include a stem and a cup of the hip implant, and the images have to be X-ray images. These images were carefully checked to avoid duplications and the clinical experts in the team evaluated each of the images to make sure that the collected X-ray images are for hip implants. For patients who had undergone total hip arthroplasty surgery, an anteroposterior (AP) view of the X-ray images for the patients with fixed (control group) and loosened hip implants were collected, while the X-ray images having a wire or plate attached with the implant were excluded. Authors managed to collect 200 hip implant X-ray images from published articles [10], [21]–[24], online resources [25]–[28], and Radiopaedia [29] and are also available the complete dataset [30]. Since these images were collected from different resources, different image resolutions, sizes, and types of implants, loosening conditions are available in this sub-set. These collected images were manually annotated by the team and finally validated by an orthopedic surgeon, who has more than 10 years of experience in THR surgery. A sample of X-ray images and corresponding implant annotations are shown in Figure 2.

2) ASEPTIC LOOSENING DETECTION SUB-SET

The original X-ray images (with single leg or dual legs) were cropped to get the hip implant section (as shown in Figure 2(i)(B)). There are 206 X-ray images with a single hip implant is available in this dataset, where hip implant images of both loose and control groups are included. The hip implant images in this database had varying resolutions (256 to 1024 pixels). Out of 206 hip implant X-ray images in the database, 112 images were from aseptic loosening patients and 94 images were from control participants. Figure 3 shows the example X-ray images of the implant detection dataset.

3) PREPROCESSING

In this study, two different experiments used different types of deep learning models with different input image size requirements, and therefore the datasets were preprocessed to resize the original X-Ray images (Implant Detection Sub-set). The state-of-the-art object detection network, YOLOv5 [31] was used for the implant detection task (first experiment), where the input image size is resized to 640 $\times$ 640 pixels. For the second experiment, popular pre-trained CNN models, such as InceptionV3 [32]–[34], ResNet [35], DenseNet [36], MobileNetV2 [37], and GoogleNet [38] were used. The X-ray images were resized to 299 $\times$ 299 for InceptionV3 and 224 $\times$ 224 for other CNN models. The images were normalized using Z-score normalization [39] with the mean and standard deviation of the entire image dataset. The pre-processing also involved the dataset preparation for the machine learning experiments which are mentioned below.

4) IMAGE AUGMENTATION AND TRAINING PARAMETERS

For five-fold cross-validation, the entire image set was divided into 80 percent training and 20 percent testing subsets, with 10% of the training dataset used for validation, with the primary goal of avoiding overfitting. The training dataset has to be balanced to avoid biased training which was done with the help of the data augmentation approach, an effective method to provide reliable results evident in many of the authors' recent publications [40]–[45]. Moreover, the image dataset is small for training deep learning models. To balance the training image classes and to make the training set larger to avoid over-fitting [46], three popular image augmentation techniques (rotation, scaling, and translation) were used. The images were rotated in a clockwise and counterclockwise direction with an angle of 5 to 10 degrees for image augmentation. The scaling operation is the magnification or reduction of the frame size of the image and 2.5% to 10% image magnifications were used in this work. Image translation was done by translating images horizontally and vertically by 5% to 10%. The number of training, validation, and test images used in implant localization and hip implant loosening detection experiments are shown in Table 1.

Details of the training parameters for implant localization and hip implant loosening detection experiments can

TABLE 1. Number of images per class and per fold used for implant localization and hip implant loosening detection.

Experiments		*Total Samples*	*Augmented Training Samples/Fold*	*Validation Samples/Fold*	*Testing Samples/ Fold*
Implant Localization		200	144×20= 2880	16	40
Loosening Detection	Control	94	67×30= 2010	8	19
	Loose	112	80×25= 2000	9	23

TABLE 2. Training parameters for implant localization and hip implant loosening detection.

	Batch size	Learning rate	Epochs	Epoch patience	Stopping epochs	Optimizer
Implant Localization	32	0.0001	300	30	25	SGD
Loosening Detection	32	0.001	30	15	15	ADAM

be seen in Table 2. The final receiver operating characteristic (ROC) curve, confusion matrix, and evaluation matrices were obtained by averaging the findings of five-fold cross-validation.

B. MACHINE LEARNING MODELS

Two different experiments of this study used two different machine learning modalities, which are explained below:

1) IMPLANT LOCALIZATION

In this study, we used a different version of YOLOv5 object detection models to detect the implant region from the hip implant X-ray images. YOLO is a state-of-the-art, real-time object detector, and Yolov5 is developed as a continuous improvement effort from Yolov1 to Yolov4 to achieve top performances on two official object detection datasets: Pascal VOC (visual object classes) [47] and Microsoft COCO (common objects in context) [48]. There are three reasons to choose Yolov5 (object detector) as the implant localizer model.

Firstly, Yolov5 incorporated a cross-stage partial network (CSPNet) [49] into Darknet, creating CSPDarknet as its backbone. CSPNet solves the problems of repeated gradient information in large-scale backbones and integrates the gradient changes into the feature map, thereby decreasing the parameters and FLOPS (floating-point operations per second) of the model, which not only ensures the inference speed and performance but also reduces the model size. Secondly, the Yolov5 applied path aggregation network (PANet) [50] as its neck to boost information flow. PANet adopts a new feature pyramid network (FPN) structure with an enhanced bottom-up path, which improves the propagation of low-level features. At the same time, adaptive feature pooling, which links the feature grid and all feature levels, is used to make useful information in each feature level propagate directly to the following subnetwork. PANet improves the utilization

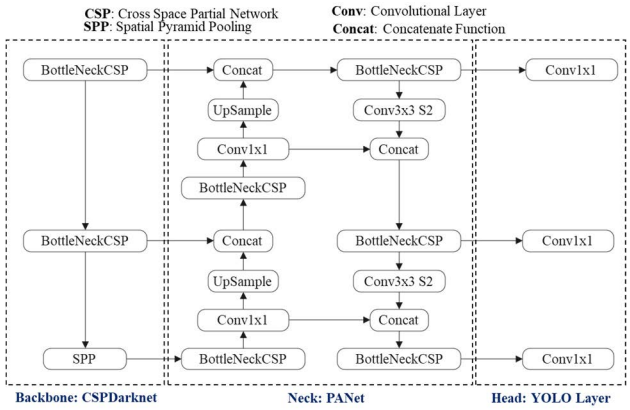

FIGURE 4. Yolov5's network architecture consists of three sections: CSPDarknet as the backbone, PANet as the neck, and Yolo Layer as the head.

of accurate localization signals in lower layers, which can enhance the location accuracy of the object. Thirdly, the head of Yolov5, namely the Yolo layer, generates 3 different sizes (18×18, 36×36, 72×72) of feature maps to achieve multi-scale [51] prediction, enabling the model to handle small, medium, and large objects efficiently. The network architecture of Yolov5 is shown in Figure 4.

In the network, the data is first supplied into CSPDarknet, which extracts features, and then into PANet, which fuses them. Finally, Yolo Layer outputs detection results (class, score, location, size). According to YOLOv5, the confidence score replicates whether a target object exists in a cell or not. Also, it predicts the object accurately. The confidence score is calculated by using the following Equation (1):

$$\text{confidence score} = P(T_o) \times IoU(G, B) \tag{1}$$

where $P(T_o)$ is the prediction of the target, T_o and prediction of the target will be in the range of (0,1); intersection over union (IoU) is calculated between G and B, where G is ground truth and B is the predicted box. The confidence score of each class is projected by the leaky rectilinear unit (ReLu) and sigmoid activation functions, and a threshold value identifies the object. In YOLOv5, the Binary Cross-Entropy with Logistic Loss (BCELL) function from the PyTorch library is used for calculating the loss of class probability and target object scores [52]. Moreover, an image contains multiple target objects, and the objects might be of different shapes and sizes. So, the target objects might be captured perfectly with a single bounding box.

Algorithm 1 : Non-Max Suppression

```
procedure NMS (B, c)
    B_nms ← ∅ Initialize empty set
    for b_i ∈ B do => Itarate over all the boxes
        discard ← False
        for b_j ∈ B do =>
Start another loop to compare with b (i)
            if the same (b_i, b_j) > γ_nms then
If both boxes having same IOU
                if score (c, b_j) > score (c, b_i) then
Compare the scores
                    discard ← True
        if not discard then
            B_nms ← B_nms ∪ b_i
    return B_nms
```

The YOLOv5 object detection model creates more than one overlapping bounding box (BB) in a single image to detect target objects but needs to show only a single bounding box for each object in an image. Thus, the Non-Maximum Suppression (NMS) technique is applied to eliminate the overlapping problem, which selects a single BB out of more than one overlapping BB to identify the objects in an image. The NMS method removes the redundant identifications and determines the best match for ending identification. The NMS technique is presented in Algorithm 1.

2) ASEPTIC LOOSENING DETECTION

In this study, we used a stacking approach for hip implant loosening detection using implant X-ray images where the eight state-of-the-art CNN models such as i) Resnet18 [34], ii) Resnet50 [34], iii) Resnet101 [34], iv) InceptionV3 [34], v) DenseNet161 [53], vi) DenseNet201 [53], vii) Mobilenetv2 [37], and viii) Googlenet [54] was investigated. Then the stacking approach was deployed with the top-performing three models as base learners and the predictions of these models were used to train ten different machine learning classifiers as meta learners to make the final decision.

If a single dataset A, which consists of input vectors ($\boldsymbol{x_i}$) and their classification score ($\boldsymbol{y_i}$). At first, a set of base-level classifiers $\boldsymbol{M_1}, \ldots\ldots, \boldsymbol{M_p}$ is trained and the prediction of these base learners is used to train the meta-level classifier $\boldsymbol{M_f}$ as illustrated in Figure 5.

We used five-fold cross-validation to generate a training set for the meta-level classifier. Among these folds, base-level classifiers were used on four folds, leaving one fold for testing. Each base-level classifier produces a probability value for the possible classes. Thus, using input x, a probability distribution is created using the predictions of the base-level classifier set, M:

$$\mathbf{P^M}(\mathbf{x}) = \left(\mathbf{P^M}(\mathbf{c_1} \mid \mathbf{x}), \mathbf{P^M}(\mathbf{c_2} \mid \mathbf{x}), \ldots\ldots, \mathbf{P^M}(\mathbf{c_n} \mid \mathbf{x})\right) \tag{2}$$

where $(\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots\ldots, \boldsymbol{c_n})$ is the set of possible class values n, m denotes the number of subjects and $\boldsymbol{P^M}(\boldsymbol{c_i} \mid \boldsymbol{x})$ denotes the probability that example, x belongs to a class $\boldsymbol{c_j}$ as estimated (and predicted) by the classifier, $\boldsymbol{M}$ in Equation (2). The class $\boldsymbol{c_i}$ with the highest-class probability $\boldsymbol{P^{M_j}}(\boldsymbol{c_i} \mid \boldsymbol{x})$ is predicted by a classifier, $\boldsymbol{M}$. The metalevel classifier $\boldsymbol{M_f}$, and attributes are thus the probabilities predicted for each possible class by each of the base-level classifiers, i.e., $\boldsymbol{P^{M_j}}(\boldsymbol{c_i} \mid \boldsymbol{x})$ for i = 1, ..., n and j = 1, ..., p where n, p denotes the number of classes and the number of base learners. The pseudo-code for the stacking approach is shown in Algorithm 2.

Algorithm 2 : Stacking Technique

Input: training data $\boldsymbol{A} = \{\boldsymbol{x_i}, \boldsymbol{y_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{m}}\boldsymbol{c}$
Output: a stacking classifier $\boldsymbol{M_f}$

```
Step 1: learn base-level classifiers
for t=1 to T do
    learn h_t based on A
end for
Step 2: construct new data set of predictions
for i =1 to m do
    A_h = {x'_i, y_i}, where x'_i = {h_1 (x_i) , . . . .., h_T (x_i)
end for
Step 3: learn a meta-classifier
learn M_f based on A_h
return M_f
```

C. PERFORMANCE METRICS

1) IMPLANT LOCALIZATION

The performance of the YOLOv5 models in localizing the implant area in the hip implant X-ray images is evaluated by the different evaluation metrics such as (i) Precision (P), (ii) Recall (R), and (iii) Mean average precision (mAP). Precision represents the ability of a model to detect only the relative objects. On the other hand, recall represents the ability of a model to find out all the relevant cases. The mAP for object detection is the mean of the average precision calculated for all the classes. The intersection over union (IoU) is given by the ratio of the area of intersection and the area of the union of the predicted bounding box and ground truth bounding box. Traditionally mAP at IoU = 0.5 is used to measure the object detection performance while mAP at IoU = 0.5-0.95 is a good performance metric. Here, mAP0.5-0.95 represents the mean average starting at IoU = 0.5 and stepping in 0.05 up to IoU = 0.95. As a result, ten distinct IoUs were calculated to compute the AP threshold. The average is used to provide a single value that rewards better localization detection.

The evaluation matrices are calculated by using the following Equations (3-5):

$$\text{Recall (R)} = \frac{TP}{TP + FN} \tag{3}$$

$$Precision\ (P) = \frac{TP}{TP + FP} \tag{4}$$

$$\text{mAP} = \sum_{I=1}^{m} P(i) \times \Delta R(i) \tag{5}$$

where True Positive (TP) is the number of correctly identified implant regions, False Positive (FP) is the number of

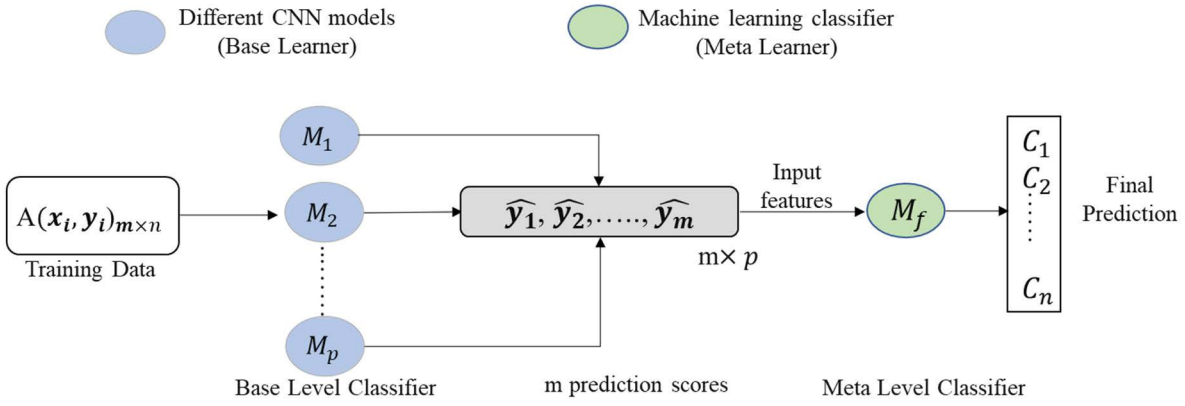

FIGURE 5. Stacking model architecture.

identified implant regions that are not implanted, and False Negative (FN) is the number of implant regions that are incorrectly identified. P(i) is the precision, and $\Delta R(i)$ is the change in recall from the ith detection.

2) ASEPTIC LOOSENING DETECTION

The performance of different CNN models and machine learning classifiers was evaluated using five performance metrics: Overall accuracy, weighted precision, weighted sensitivity or recall, weighted F1-score, and weighted specificity using Equations (6-10). As various classes have varying numbers of images, the networks were compared using a per-class weighted performance metric and overall accuracy. The area under the curve (AUC) was also used to assess the performance.

$$Accuracy = \frac{(TP + TN)}{(TP + FN) + (FP + TN)} \tag{6}$$

$$Precision = \frac{(TP)}{(TP + FP)} \tag{7}$$

$$Sensitivity = \frac{(TP)}{(TP + FN)} \tag{8}$$

$$F1{-}Score = \frac{(2 * TP)}{(2 * TP + FN + FP)} \tag{9}$$

$$Specificity = \frac{(TN)}{(FP + TN)} \tag{10}$$

Here, true positive (TP), true negative (TN), false positive (FP), and false-negative (FN) were used to denote the number of loose hip implant X-ray images were identified as loose, the number of control group hip implant X-rays were identified as control, the number of control group hip implant X-rays incorrectly identified as loose and the number of loose hip implant X-ray images incorrectly identified as control, respectively. We report the weighted performance metric, with a 95 % confidence interval, for Sensitivity, Specificity, Precision, and F1-Score, and the overall accuracy with a 95 % confidence interval for the accuracy.

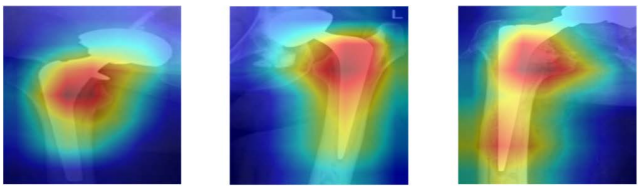

FIGURE 6. Score-CAM visualization of some loose hip implant X-ray images, to demonstrate the region where the CNN model is mostly making decisions from.

D. VISUALIZATION TECHNIQUES

With the emergence of visualization tools, there has been a rise in curiosity about how CNN works and the reasoning underlying its decision-making. Visualization approaches improve the visual portrayal of the decision-making process of CNNs. These also improve the model's transparency by showing the rationale behind the inference in a way that humans can understand, hence enhancing trust in the CNNs' outputs. Score-CAM was chosen for this investigation because of its promising performance in recent computer vision medical problems [18], [55]. Figure 6 shows a Score-CAM visualization that highlights the regions that CNN considers when making decisions. By confirming decision-making from important regions of the images, these visualizations serve to increase trust in the reliability of deep layer networks.

III. RESULTS

This section discusses the results of the implant localization and aseptic loosening detection experiments along with the Score-CAM visualization to better interpret the model performance.

A. IMPLANT LOCALIZATION

We investigated different versions of YoloV5 object detection models such as YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x to detect implant regions from hip implant X-ray images. Firstly, the precision, recall and mean average precision performance metrics with respect to epochs of four

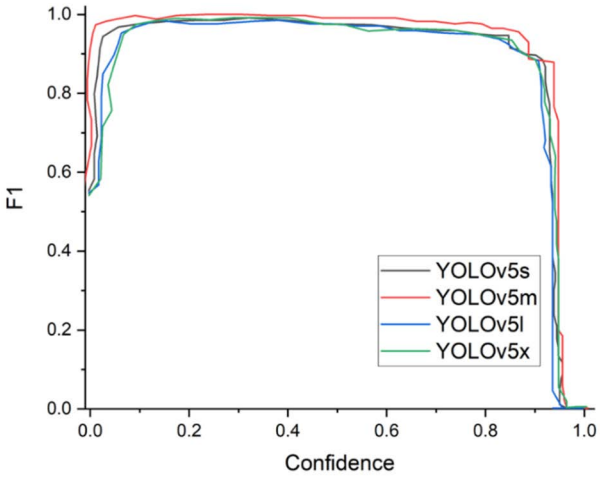

FIGURE 7. F1 score performance curve for implant detection from hip implant X-ray image with respect to confidence score using different versions of the YOLOv5 model.

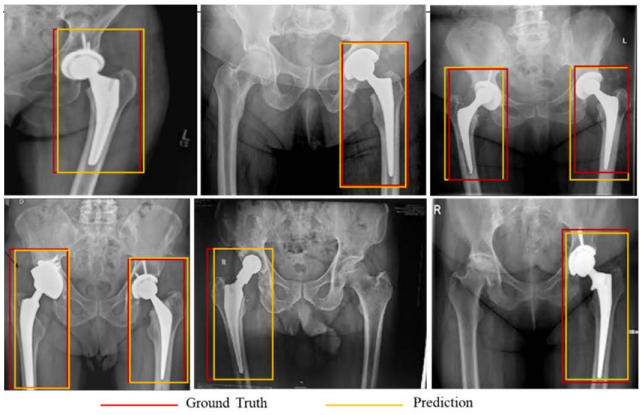

FIGURE 8. Sample test images using the best performing implant localization model, YOLOv5m.

versions of the YOLOv5 model are investigated. Precision is the ability of a model to detect only the implant objects, whereas recall or sensitivity measures the number of objects correctly detect as an implant. We also investigated the mean average precision performance with IoU = 0.5, which is a traditional performance metric to measure the performance of object detection models, as well as investigated mAP with IoU 0.5 to 0.95. In this study, all versions of the YOLOv5 object detection models performed very well with 100% precision, recall, and mAP0.5. Table 3 shows the comparison of different versions of YOLOv5 model performances for implant detection using hip implant X-ray images.

Although all versions of YOLOv5 object detection models performed with 100% precision and recall, using mAP0.5-0.95 it is apparent that YOLOv5m performed slightly better than other versions of YOLOv5 (YOLOv5s, YOLOv5l, and YOLOv5x). The mean average precision mAP0.5-0.95 for YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x is 86.5%, 87.8%, 86.5%, and 84.9% respectively. The F1 score performance curve is depicted in Figure 7. The F1 score is calculated based on precision and recall, where it shows that F1 scores performance curve for implant detection from hip implant X-ray image with respect to confidence score using different versions of the YOLOv5 model. It is also evident in Figure 7 that the YOLOv5m version outperformed other versions of YOLOv5 in terms of implant detection using hip implant X-ray images.

Figure 8 shows that different versions of YOLOv5 models trained on the hip implant X-ray dataset can detect the implant areas of the X-ray images very reliably. However, the YOLOv5m was used for the remaining experiment as it outperformed other models. Some sample test images with ground truth bounding boxes and predicted bounding boxes are shown in Figure 8, where it is clearly seen that ground truth and predicted bounding boxes for implant detection are almost overlapping each other.

B. ASEPTIC LOOSENING DETECTION

This section describes the performance of the different classification networks in detecting control and loose hip implant X-ray images. As mentioned earlier, eight different state-of-the-art CNN networks and a stacking approach with the top-performing networks were investigated to identify loosening of the hip implants using X-ray images. The comparative performance of different CNNs for these classification schemes is shown in Table 2 (A).

The best classification accuracy, precision, sensitivity, F1 score, and specificity for loosening detection were found to be 94.66%, 94.66%, 94.66%, 94.66%, and 94.5%, respectively using the Densenet201 CNN model.

For binary classification (control vs loosening) using hip implant X-ray images, the top-performing three CNN models were DenseNet201, Resnet50, Resnet18 with an overall accuracy of 94.66%, 93.69%, and 91.75%, respectively. We used these three top-performing models as base learners in the stacking approach, where the predictions of these three models were used as input to another meta learner. Ten different machine learning classifiers were investigated as meta learners and we found Random Forest classifier outperformed other classifiers with accuracy, precision, sensitivity, F-1 score, and specificity of 96.11%, 96.42%, 96.42%, 96.42%, and 96.74%, respectively for loosening detection (Table 4(B)).

Figure 9 shows the area under the curve (AUC)/receiver-operating characteristics (ROC) curve (also known as AUROC (area under the receiver operating characteristics)) for loosening detection using hip implant X-ray images, which is one of the most important evaluation metrics for checking any CNN model's performance. This is apparent from the ROC curves that the DenseNet201 CNN model outperformed other networks for classification with 97.68% AUC in Figure 9(A). In the stacking model, Random Forest classifiers were the best performer as meta learners with 98.94% AUC which is shown in Figure 9(B).

TABLE 3. Comparison of different versions of YOLOv5 model performances for implant detection using hip implant X-ray images.

Networks	Precision	Recall	$mAP_{0.5}$	$mAP_{0.5-0.95}$	Inference time (sec)	Number of parameters (in millions)	Number of layers
YOLOv5s	100	100	100	86.5	0.01	7	283
YOLOv5m	**100**	**100**	**100**	**87.8**	**0.017**	**21.1**	**391**
YOLOv5l	100	100	100	86.5	0.022	46.6	499
YOLOv5x	100	100	100	84.9	0.03	88	672

TABLE 4. Comparison of different CNN performances for binary classification for (A) different CNN models and (B) stacking machine learning classifiers (A) different CNN models and (B) stacking machine learning classifiers.

(A)

Networks	Overall Accuracy	Weighted average with 95% CI				
		Precision	Sensitivity	F1-score	Specificity	Inference time (sec)
Resnet18	**91.75 ± 3.76**	**91.75 ± 3.76**	**91.75 ± 3.76**	**91.75 ± 3.76**	**91.54 ± 3.8**	**0.2**
Resnet50	**93.69 ± 3.32**	**93.7 ± 3.32**	**93.69 ± 3.32**	**93.69 ± 3.32**	**93.68 ± 3.32**	**0.35**
Resnet101	91.26 ± 3.86	91.3 ± 3.85	91.26 ± 3.86	91.25 ± 3.86	90.79 ± 3.95	0.65
InceptionV3	90.29 ± 4.04	90.39 ± 4.02	90.29 ± 4.04	90.31 ± 4.04	90.49 ± 4.01	1.02
DenseNet161	90.29 ± 4.04	90.32 ± 4.04	90.29 ± 4.04	90.3 ± 4.04	90.32 ± 4.04	048
DenseNet201	**94.66 ± 3.07**	**94.66 ± 3.07**	**94.66 ± 3.07**	**94.66 ± 3.07**	**94.5 ± 3.11**	**0.6**
Mobilenetv2	89.81 ± 4.13	89.8 ± 4.13	89.81 ± 4.13	89.8 ± 4.13	89.56 ± 4.18	0.13
GoogleNet	91.26 ± 3.86	91.26 ± 3.86	91.26 ± 3.86	91.26 ± 3.86	90.96 ± 3.92	0.22

(B)

Classifiers	Overall Accuracy	Weighted average with 95% CI				
		Precision	Sensitivity	F1-score	Specificity	Inference time (sec)
Multi-layer perceptron (MLP)	94.66 ± 3.07	94.66 ± 3.07	94.66 ± 3.07	94.66 ± 3.07	94.66 ± 3.07	0.381
Linear Discriminant Analysis	94.17 ± 3.2	94.17 ± 3.2	94.17 ± 3.2	94.17 ± 3.2	94.17 ± 3.2	0.016
XGBoost	92.72 ± 3.55	92.72 ± 3.55	92.72 ± 3.55	92.73 ± 3.55	92.72 ± 3.55	0.068
Extra Tree	95.55 ± 3.12	95.56 ± 3.11	95.55 ± 3.12	95.55 ± 3.13	95.6 ± 3.11	0.166
Logistic Regression	93.69 ± 3.32	93.69 ± 3.32	93.69 ± 3.32	93.69 ± 3.32	93.69 ± 3.32	0.085
Support Vector Machine (SVM)	91.26 ± 3.86	91.26 ± 3.86	91.26 ± 3.86	91.28 ± 3.85	91.26 ± 3.86	0.016
Random Forest	**96.11 ± 2.64**	**96.42 ± 2.54**	**96.42 ± 2.54**	**96.42 ± 2.54**	**96.74 ± 2.74**	**0.112**
AdaBoost	94.17 ± 3.2	94.17 ± 3.2	94.17 ± 3.2	94.18 ± 3.2	94.17 ± 3.2	0.09
K-Nearest Neighbors (KNN)	94.66 ± 3.07	94.66 ± 3.07	94.66 ± 3.07	94.66 ± 3.07	94.66 ± 3.07	0.06
Gradient Boosting	95.63 ± 2.79	95.63 ± 2.79	95.63 ± 2.79	95.64 ± 2.79	95.63 ± 2.79	0.054

Figure 10 shows the confusion matrix for the best performing CNN model and the stacking model with Random Forest meta learner for loosening detection using hip implant X-ray images.

Figure 10(A) shows the confusion matrix of the best performing CNN model (DenseNet201) and Figure 10(B) shows the confusion matrix of the best performing stacking CNN model (with Random Forest classifier as a meta learner). The best performing DenseNet201 network failed to detect 5 out of 112 loose hip implant X-ray images while incorrectly detected 6 control group hip implant X-ray images as loose whereas 108 out of 112 loose hip implant X-ray images were correctly detected as loose and 90 out of 94 control hip implant X-ray images were correctly identified as control

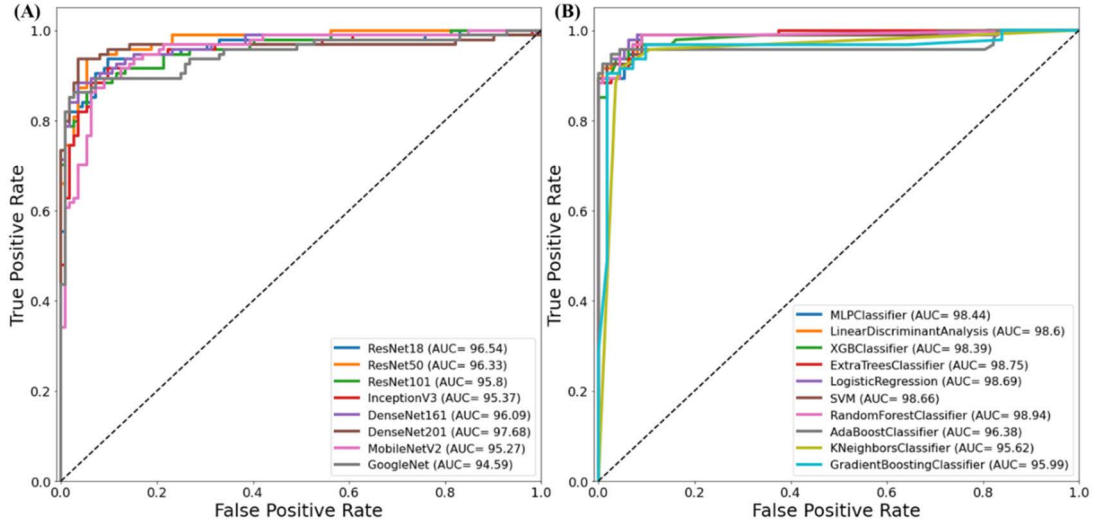

FIGURE 9. ROC curve for Control and Loose hip implant classification using (A) different CNN networks and (B) stacking meta learner classifiers.

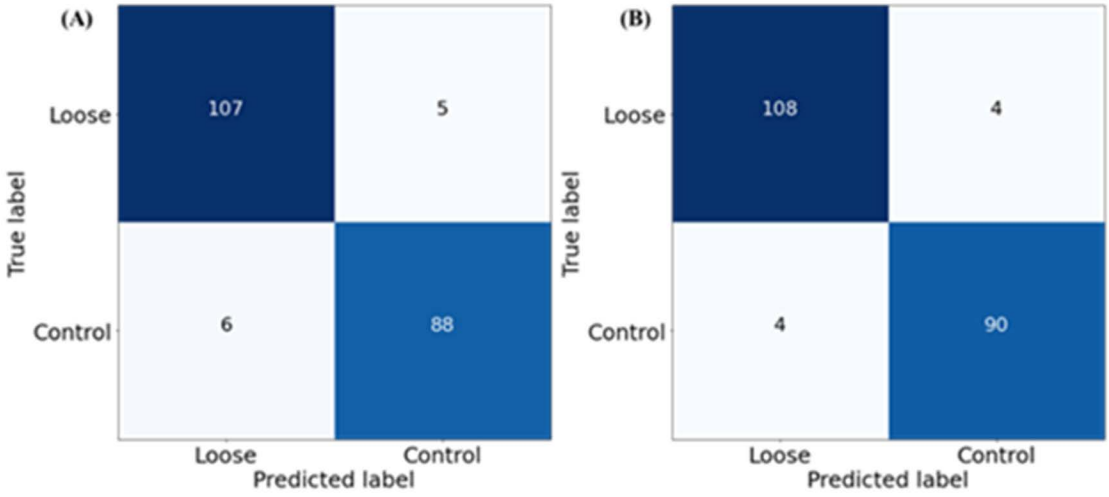

FIGURE 10. Confusion matrix for Control and Loose hip implant classification using (A) the best performing CNN model, (B) the best performing stacking CNN model (with Random Forest meta learner).

images with stacking CNN model. Thus, this is evident that the stacking CNN model with Random Forest classifier as a meta learner outperformed other state-of-the-art CNN models.

IV. DISCUSSION

The study carried out two major experiments: i) four different object detection models were investigated in detecting the implant from the hip implant X-ray images and ii) eight different CNN models and stacking models were investigated to classify the loosening in the hip implant X-ray images. The performance of the YOLOv5m model exceeded other models in detecting the implant region from hip implant X-ray images, with precision, recall, $mAP_{0.5}$, $mAP_{0.5-0.95}$ of 100%, 100%, 100%, and 87.8%, respectively. For different CNN models, the Densenet201 model outperformed others with the accuracy, precision, sensitivity, F1-score, and specificity of 94.66%, 94.66%, 94.66%, 94.66%, and 94.5%, respectively. However, the stacking CNN approach with Random Forest meta learner classifier produced the best performance with the accuracy, precision, sensitivity, F1-score, and specificity of 96.11%, 96.42%, 96.42%, 96.42%, and 96.74%, respectively for loosening detection. Moreover, it was also confirmed that the stacking approach can improve the detection accuracy by around 2%.

To the best of the author's knowledge, this kind of extensive investigation was not done before for hip or knee implant loosening detection and the authors have compared this state-of-the-art performance with similar works on the hip implant (but on a very small dataset) and on the knee implant in Table 5. In a previous study [56], the loosening classification sensitivity has been reported to be 94% but they used a

TABLE 5. Comparison of different CNN performances for binary classification for (A) different CNN models and (B) stacking machine learning classifiers comparison of the findings of this study with recent similar studies.

Author	Year	Method	Database	Results
Borjali et al. [58]	2019	Deep convolutional neural network	40 X-rays	Specificity – 94% and Sensitivity – 94 %
Shah et al. [59]	2020	Deep convolutional neural network	697 X-rays	Accuracy, sensitivity, and specificity – 88.3%, 70.2%, and 95.6%
This Study	2022	Implant detection using different versions of YOLOv5	200 X-rays for implant localization	For implant localization, Precision-100%, $mAP_{0.5}$-100%, $mAP_{0.5-0.95}$-87.8%
		A stacking approach using different CNN models for loosening detection	206 X-rays for loosening detection	For loosening detection, Specificity – 96.74% and Sensitivity – 96.42 %

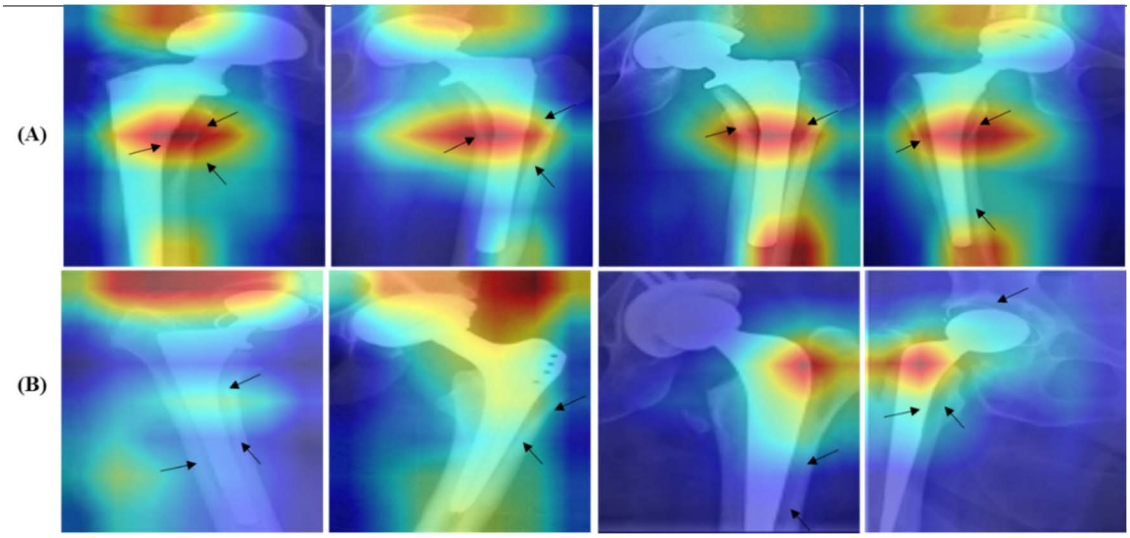

FIGURE 11. Score-CAM visualization of (A) correctly classified, and (B) misclassified loose hip implant X-ray images. Note: Black arrows indicate the loosening regions of the hip implant X-ray images.

database consisting of a small number (only 40) of X-ray images. In [57], a larger private dataset was used for knee implant loosening detection without object detection, where the reported model has shown 88.3% accuracy. However, in this study, a dataset of 200 X-ray images (labeled by 10+ years experienced as a radiologist) was used and a better result is obtained compared to both studies. Moreover, the datasets used by the other studies were not made public to evaluate the performance of our model on that dataset while we have made our dataset public so that interested researchers can replicate our results easily. In addition, the implant localization approach has helped to improve loosening classification/detection performance with a novel stacking CNN model, making the complete approach more robust and versatile, with a detection sensitivity of 96.42 percent.

Score-CAM-based heat maps were generated for X-ray images to see the saliency maps of the network's prediction. Figure 11(A) depicts the Score-CAM visualization of correctly identified loose hip implant X-ray images, while Figure 11(B) depicts the same for misclassified loose hip implant X-rays using the best performing model. Moreover, it is visible that the model is deciding on the loosening area of the hip implant X-ray and black arrows indicate the loosening region in the hip implant X-ray images. It is noticed that for the misclassified images, the model is taking a decision from the non-relevant areas (loosening) and it is also visible that most of the misclassified images are in the early stage of loosening. This heat map or saliency map increases end-user confidence in the network's output, which makes the deep learning model more explainable and reliable as a computer-aided diagnostic tool.

V. CONCLUSION

This study used a stacking approach with deep Convolutional Neural Networks to detect aseptic loosening in hip implant radiographs automatically. Moreover, this research

looked into two main important experiments. Firstly, the performance of four different object detection models was evaluated to localize the implant region from the hip implant X-ray images. YOLOv5m model outperformed other models to detect the implant region from the hip implant X-ray images with the precision, recall, mAP0.5, mAP0.5-0.95 of 100%, 100%, 100%, and 87.8% respectively. Secondly, the performance of stacking models was evaluated to classify the loosening in the hip implant X-ray images and compared with eight different state-of-the-art CNN models. Densenet201 CNN model outperformed other CNN models with the accuracy, precision, sensitivity, F1-score, and specificity of 94.66%, 94.66%, 94.66%, 94.66%, and 94.5%, respectively whereas the stacking CNN approach with Random Forest meta learner classifier produced the best performance with the accuracy, precision, sensitivity, F1-score and specificity of 96.11%, 96.42%, 96.42%, 96.42%, and 96.74%, respectively for loosening detection. It was also confirmed that the stacking approach can improve the detection accuracy by around 2%. The Score-CAM visualization output demonstrates that the loosening detection decision of the model was done based on the relevant region of the hip implant X-ray images when the model correctly detect the loosening. The performance of the study could be improved with a larger dataset to increase the robustness of the model, but as mentioned earlier publically available dataset is not accessible, thus the future study would try to have a larger dataset with severity labeling (mild, moderate, and severe aseptic loosening hip implant X-ray) by the experts, which can help in following-up the disease progression and allow clinicians to apply the intervention to delay the revision surgery requirement. It can be concluded that the newly developed AI-based aseptic loosening system can aid in diagnosis in the presence/absence of an expert radiologist, which can help in reducing a large number of revision surgeries, which might have happened due to delayed or improper diagnosis.

ACKNOWLEDGMENT

The statements made herein are solely the responsibility of the authors.

DATA AVAILABILITY

The database [30] created as a part of this study is made publicly available for the researchers to take benefit from this study. Once the article got accepted, the dataset will be made public.

ETHICAL CONSIDERATION

Since the dataset was created using the publicly available images and the authors did not collect them from the hospital, there is no ethical approval is required for creating this dataset. Moreover, the user identity is not available in the X-ray images and the dataset is entirely de-identified and anonymous

REFERENCES

[1] M. Ribeiro, F. J. Monteiro, and M. P. Ferraz, "Infection of orthopedic implants with emphasis on bacterial adhesion process and techniques used in studying bacterial-material interactions," *Biomatter*, vol. 2, no. 4, pp. 176–194, Oct. 2012.
[2] H. A. Ching, D. Choudhury, M. J. Nine, and N. A. A. Osman, "Effects of surface coating on reducing friction and wear of orthopaedic implants," *Sci. Technol. Adv. Mater.*, vol. 15, no. 1, Feb. 2014, Art. no. 014402.
[3] W. Huang, E. Zalnezhad, F. Musharavati, and P. Jahanshahi, "Investigation of the tribological and biomechanical properties of CrAlTiN and CrN/NbN coatings on SST 304," *Ceram. Int.*, vol. 43, pp. 7992–8003, 2017.
[4] H. D. Homesley, J. M. Minnich, J. Parvizi, and W. J. Hozack, "Total hip arthroplasty revision: A decade of change," *Amer. J. Orthopedics*, vol. 33, no. 8, pp. 389–392, 2004.
[5] M. R. Dayton and S. J. Incavo, "Component loosening in total hip arthroplasty," *Seminars Arthroplasty*, vol. 16, no. 2, pp. 161–170, Jun. 2005.
[6] R. Mulroy and W. Harris, "The effect of improved cementing techniques on component loosening in total hip replacement. An 11-year radiographic review," *J. Bone Joint Surg. Brit. Volume*, vols. 72–B, no. 5, pp. 757–760, Sep. 1990.
[7] K. S. Keisu, F. Orozco, J. D. McCallum, G. Bissett, W. J. Hozack, P. F. Sharkey, and R. H. Rothman, "Cementless femoral fixation in the rheumatoid patient undergoing total hip arthroplasty: Minimum 5-year results," *J. Arthroplasty*, vol. 16, no. 4, pp. 415–421, Jun. 2001.
[8] B. G. Evans, E. A. Salvati, M. H. Huo, and O. L. Huk, "The rationale for cemented total hip arthroplasty," *Orthopedic Clinics North Amer.*, vol. 24, pp. 599–610, Oct. 1993.
[9] A. Huang, W. Palmer, and C. Chang, "Radiographic evaluation of hip implants," *Seminars Musculoskeletal Radiol.*, vol. 19, no. 1, pp. 12–20, Jan. 2015.
[10] J. Vanrusselt, M. Vansevenant, G. Vanderschueren, and F. Vanhoenacker, "Postoperative radiograph of the hip arthroplasty: What the radiologist should know," *Insights Imag.*, vol. 6, no. 6, pp. 591–600, Dec. 2015.
[11] O. P. P. Temmerman, P. G. H. M. Raijmakers, J. Berkhof, E. F. L. David, R. Pijpers, M. A. Molenaar, O. S. Hoekstra, G. J. J. Teule, and I. C. Heyligers, "Diagnostic accuracy and interobserver variability of plain radiography, subtraction arthrography, nuclear arthrography, and bone scintigraphy in the assessment of aseptic femoral component loosening," *Arch. Orthopaedic Trauma Surgery*, vol. 126, no. 5, pp. 316–323, Jul. 2006.
[12] O. P. P. Temmerman, P. G. H. M. Raijmakers, E. F. L. David, R. Pijpers, M. A. Molenaar, O. S. Hoekstra, J. Berkhof, R. A. Manoliu, G. J. J. Teule, and I. C. Heyligers, "A comparison of radiographic and scintigraphic techniques to assess aseptic loosening of the acetabular component in a total hip replacement," *J. Bone Joint Surg.*, vol. 86, no. 11, pp. 2456–2463, 2004.
[13] C. Khalily and L. A. Whiteside, "Predictive value of early radiographic findings in cementless total hip arthroplasty femoral components: An 8-to 12-year follow-up," *J. Arthroplasty*, vol. 13, pp. 768–773, Oct. 1998.
[14] T. O. Smith, T. H. D. Williams, A. Samuel, L. Ogonda, and J. A. Wimhurst, "Reliability of the radiological assessments of radiolucency and loosening in total hip arthroplasty using PACS," *HIP Int.*, vol. 21, no. 5, pp. 577–582, Sep. 2011.
[15] A. M. Tahir, Q. Yazan, A. Khandakar, T. Rahman, U. Khurshid, F. Musharavati, M. T. Islam, S. Kiranyaz, S. Al-Maadeed, and M. E. H. Chowdhury, "Deep learning for reliable classification of COVID-19, MERS, and SARS from chest X-ray images," *Cogn. Comput.*, pp. 1–21, 2022, doi: 10.1007/s12559-021-09955-1.
[16] Y. Qiblawey, A. Tahir, M. E. H. Chowdhury, A. Khandakar, S. Kiranyaz, T. Rahman, N. Ibtehaz, S. Mahmud, S. A. Maadeed, F. Musharavati, and M. A. Ayari, "Detection and severity classification of COVID-19 in CT images using deep learning," *Diagnostics*, vol. 11, no. 5, p. 893, May 2021.
[17] A. Khandakar, M. E. H. Chowdhury, M. B. I. Reaz, S. H. M. Ali, M. A. Hasan, S. Kiranyaz, T. Rahman, R. Alfkey, A. A. A. Bakar, and R. A. Malik, "A machine learning model for early detection of diabetic foot using thermogram images," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104838.
[18] T. Rahman, A. Khandakar, M. A. Kadir, K. R. Islam, K. F. Islam, R. Mazhar, T. Hamid, M. T. Islam, S. Kashem, Z. B. Mahbub, M. A. Ayari, and M. E. H. Chowdhury, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020.
[19] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Rev. Cancer*, vol. 18, no. 8, pp. 500–510, 2018.

[20] G. Jocher. (2022). *Ultralytics/YOLOv5*. [Online]. Available: https://github.com/ultralytics/yolov5
[21] T. Ilchmann, W. Zimmerli, P. E. Ochsner, B. Kessler, L. Zwicky, P. Graber, and M. Clauss, "One-stage revision of infected hip arthroplasty: Outcome of 39 consecutive hips," *Int. Orthopaedics*, vol. 40, no. 5, pp. 913–918, May 2016.
[22] Y.-S. Kim, C.-M. Cho, K.-T. Hwang, Y.-H. Kim, and I.-Y. Choi, "Revision total hip arthroplasty using a Wagner revision stem," *J. Korean Hip Soc.*, vol. 22, no. 2, pp. 137–142, 2010.
[23] A. Balanika, S. Theocharakis, S. Vrizidou, C. Drosos, and C. Baltas, "Radiographic interpretation of hip replacement hardware. A pictorial essay," in *Proc. Eur. Congr. Radiol. (ESSR)*, 2014, doi: 10.1594/essr2014/P-0050.
[24] O. Awan, L. Chen, and C. S. Resnik, "Imaging evaluation of complications of hip arthroplasty: Review of current concepts and imaging findings," *Can. Assoc. Radiol. J.*, vol. 64, no. 4, pp. 306–313, Nov. 2013.
[25] *Radiologic Evaluation of Hip Arthroplasty*. Accessed: Apr. 4, 2021. [Online]. Available: https://plasticsurgerykey.com/radiologic-evaluation-of-hip-arthroplasty/
[26] *Revision Surgery for Hip Replacements*. Accessed: Apr. 4, 2021. [Online]. Available: http://www.cardiffhipandknee.com/hip/hip-revisions
[27] *Looseing*. Accessed: Apr. 4, 2021. [Online]. Available: http://www.amilcaregentili.com/thr/loosenin.htm
[28] *THA Aseptic Loosening*. Accessed: Apr. 5, 2021. [Online]. Available: https://www.orthobullets.com/recon/12304/tha-aseptic-loosening
[29] *Aseptic Looseing of Hip Joint Replacements*. Accessed: Jan. 1, 2021 [Online]. Available: https://radiopaedia.org/articles/aseptic-loosening-of-hip-joint-replacements?lang=us
[30] *Aspetic-Loose-Hip-Implant-Xray-Database*. Accessed: Jan. 22, 2022. [Online]. Available: https://www.kaggle.com/tawsifurrahman/aseptic-loose-hip-implant-xray-database
[31] K. Liu, H. Tang, S. He, Q. Yu, Y. Xiong, and N. Wang, "Performance validation of Yolo variants for object detection," in *Proc. Int. Conf. Bioinf. Intell. Comput.*, Jan. 2021, pp. 239–243.
[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016 pp. 2818–2826, doi: 10.1109/CVPR.2016.308.
[33] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *Proc. 2nd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2017, pp. 783–787.
[34] *ResNet, AlexNet, VGGNet, Inception: Understanding Various Architectures of Convolutional Networks*. Accessed: Dec. 23, 2020. [Online]. Available: https://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[36] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.
[37] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[39] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, Dec. 2005.
[40] C. P. Adans-Dester *et al.*, "Can mHealth technology help mitigate the effects of the COVID-19 pandemic?" *IEEE Open J. Eng. Med. Biol.*, vol. 1, pp. 243–248, 2020.
[41] A. Tahir, Y. Qiblawey, A. Khandakar, T. Rahman, U. Khurshid, F. Musharavati, M. T. Islam, S. Kiranyaz, and M. E. H. Chowdhury, "Deep learning for reliable classification of COVID-19, MERS, and SARS from chest X-ray images," 2020, *arXiv:2005.11524*.
[42] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and M. T. Islam, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132665–132676, 2020.
[43] T. Rahman, M. E. H. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Appl. Sci.*, vol. 10, no. 9, p. 3233, May 2020.
[44] M. E. H. Chowdhury, T. Rahman, A. Khandakar, S. Al-Madeed, S. M. Zughaier, H. Hassen, and M. T. Islam, "An early warning tool for predicting mortality risk of COVID-19 patients using machine learning," *Cogn. Comput.*, pp. 1–16, 2021, doi: 10.1007/s12559-020-09812-7.
[45] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. A. Maadeed, S. M. Zughaier, M. S. Khan, and M. E. H. Chowdhury, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput. Biol. Med.*, vol. 132, May 2021, Art. no. 104319.
[46] (Jul. 7, 2020). *Overfitting in Machine Learning: What it is and How to Prevent it*. [Online]. Available: https://elitedatascience.com/overfitting-in-machine-learning
[47] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2014.
[48] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
[49] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 390–391.
[50] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "PANet: Few-shot image semantic segmentation with prototype alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 9197–9206.
[51] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[52] G. Jocher. (Dec. 20, 2021). *BCEWithLogitsLoss*. [Online]. Available: https://pytorch.org/docs/master/generated/torch.nn.BCEWithLogitsLoss.html#bcewithlogitsloss
[53] *DenseNet: Better CNN Model Than ResNet*. Accessed: Dec. 29, 2020 [Online]. Available: http://www.programmersought.com/article/7780717554/
[54] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
[55] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 24–25.
[56] A. Borjali, A. F. Chen, O. K. Muratoglu, M. A. Morid, and K. M. Varadarajan, "Detecting mechanical loosening of total hip replacement implant from plain radiograph using deep convolutional neural network," 2019, *arXiv:1912.00943*.
[57] R. F. Shah, S. A. Bini, A. M. Martinez, V. Pedoia, and T. P. Vail, "Incremental inputs improve the automated detection of implant loosening using machine-learning algorithms," *Bone Joint J.*, vol. 102, pp. 101–106, Jun. 2020.

TAWSIFUR RAHMAN received the B.Sc. (Eng.) degree from the Department of Electrical and Electronic Engineering, University of Chittagong, Bangladesh, and the M.Sc. degree from the Department of Biomedical Physics and Technology (BPT), University of Dhaka, Bangladesh. He is currently working as a Research Assistant with the Department of Electrical Engineering, Qatar University. He has published several journal articles on medical imaging. His current research interests include biomedical image and signal processing, machine learning, computer vision, and data science. He has expertise in designing and developing deep CNN models using PyTorch and TensorFlow framework and has expertise in implementing nerve stimulators for measuring conduction velocity in the human body, developing electrocardiogram (ECG), electromyogram (EMG) circuit, developing Howland constant current source and instrumentation amplifier to measure tetra polar bio-impedance, and detection of the different stage of brain activity by analyzing various EEG wave. He has received an ICT Fellowship (2019–2020) from the ICT Ministry of Bangladesh for research entitled "Driver drowsiness detection from HRV and computer vision using machine learning." He and his team have recently won the COVID-19 Dataset Award for their contribution to the fight against COVID-19.

AMITH KHANDAKAR (Senior Member, IEEE) received the B.Sc. degree in electronics and telecommunication engineering from North South University, Bangladesh, and the master's degree in computing (networking concentration) from Qatar University, in 2014. He graduated as the Valedictorian (President Gold Medal Recipient) of North South University. After graduation, he was a consultant in a reputed insurance company in Qatar and in a private company that is a sub-contractor to national telecom service provider in Qatar. He is a certified Project Management Professional and the Cisco Certified Network Administrator. He was a Teaching Assistant and a Laboratory Instructor for two years for courses, such as mobile and wireless communication systems, principle of digital communications, introduction to communication, calculus and analytical geometry, and Verilog HDL: modeling, simulation, and synthesis. Simultaneously, he was a Laboratory Instructor for the following courses: programming course "C," Verilog HDL, and general physics course. He has been with Qatar University, since 2010. He is currently the General Secretary of the IEEE Qatar Section and the Qatar University IEEE Student Branch Coordinator and an Adviser (Faculty Member).

KHANDAKER REAJUL ISLAM, photograph and biography not available at the time of publication.

MD MOHIUDDIN SOLIMAN is currently pursuing the Ph.D. degree with the Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia (UKM).

MOHAMMAD TARIQUL ISLAM (Senior Member, IEEE) is currently a Professor with the Department of Electrical, Electronic, and Systems Engineering, Universiti Kebangsaan Malaysia (UKM), and a Visiting Professor with the Kyushu Institute of Technology, Japan. He is the author and coauthor of about 500 research journal articles, nearly 175 conference papers, and a few book chapters on various topics related to antennas, microwaves, and electromagnetic radiation analysis with 20 inventory patents filed. Thus far, his publications have been cited 6000 times and his H-index is 38 (Source: Scopus). His Google scholar citation is 8200 and his H-index is 42. He was a recipient of more than 40 research grants from the Malaysian Ministry of Science, Technology and Innovation, Ministry of Education, UKM Research Grant, and international research grants from Japan and Saudi Arabia. He has supervised about 30 Ph.D. thesis and 20 M.Sc. thesis, and has mentored more than ten postdocs and visiting scholars. His research interests include communication antenna design, satellite antennas, and electromagnetic radiation analysis. He is a Chartered Professional Engineer (C.Eng.); a Member of IET, U.K.; and a Senior Member of IEICE, Japan. He received several International Gold Medal awards, the Best Invention in Telecommunication Award for his research and innovation, and the Best Researcher Awards at UKM, in 2010 and 2011. He was a recipient of the 2018 IEEE AP/MTT/EMC Excellent Award. He also won the Best Innovation Award, in 2011, and the Best Research Group in the ICT niche by UKM, in 2014. He was a recipient of the Publication Award from the Malaysian Space Agency, in 2014, 2013, 2010, and 2009; and the Best Paper Presentation Award in the 2012 International Symposium on Antennas and Propagation (ISAP 2012), Nagoya, Japan, and IconSpace, in 2015. He currently serves as the Guest Editor for *Sensors* journal and an Associate Editor for IEEE Access. He was an Associate Editor of *IET Electronics Letter*.

AHMED ELSAYED is currently a Medical Expert with the Orthopaedic Department, Hamad Medical Corporation (HMC), Doha, Qatar, and Clinical Orthopaedic Surgery, Weill Cornell Medical College, Doha.

YAZAN QIBLAWEY (Member, IEEE) received the B.Sc. degree in electrical engineering from Qatar University, Qatar, in 2017, and the M.Sc. degree in electrical and electronics engineering from the University of Nottingham, U.K., in 2019. He is currently a Research Assistant with the College of Engineering, Qatar University. His research interests include signal processing, machine learning, and fault detection.

SAKIB MAHMUD is currently pursuing the master's degree with the Department of Electrical Engineering, Qatar University. His research interest includes the application of machine learning.

ASHIQUR RAHMAN received the Ph.D. degree from the Department of Electrical, Electronic, and Systems Engineering, Universiti Kebangsaan Malaysia (UKM). He is currently a Research Assistant Professor with the Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Japan.

FARAYI MUSHARAVATI received the Ph.D. degree in manufacturing systems engineering from University Putra Malaysia, in 2008. He is currently an Associate Professor with the Department of Mechanical and Industrial Engineering, Qatar University. His research interests include advanced manufacturing systems and optimization, genetic algorithms, neural networks, simulated annealing, supply chain management, and renewable energy.

ERFAN ZALNEZHAD received the Ph.D. degree from the Department of Mechanical Engineering, University of Malaya, Malaysia. He is currently with the Department of Chemical Engineering and Biomedical Engineering, The University of Texas at San Antonio, USA. His research interest includes the tribe-bio-mechanical properties investigation of bio-composites. He has published more than 90 technical papers in reputable journals and conferences. He is a reviewer for engineering ISI journals and refereed international conferences. He has conducted a lot of research on materials design, nanomaterials, thin film coatings, biomaterials, biomedical implants, biosensors, and characterization of the tribo-bio-mechanical properties of new and improved materials, including aluminum, titanium, and magnesium alloys. He is interested in the application of finite element methods and other numerical techniques and imaging and image analysis to calculate and visualize the response of biological systems to different conditions.

MUHAMMAD E. H. CHOWDHURY (Senior Member, IEEE) received the Ph.D. degree from the University of Nottingham, U.K., in 2014. He worked as a Postdoctoral Research Fellow and a Hermes Fellow at the Sir Peter Mansfield Imaging Centre, University of Nottingham. He is currently working as an Assistant Professor with the Department of Electrical Engineering, Qatar University. He has four patents and published more than 100 peer-reviewed journal articles, conference papers, and four book chapters. His current research interests include biomedical instrumentation, signal processing, wearable sensors, medical image analysis, machine learning, embedded systems design, and simultaneous EEG/fMRI. He is currently running several NPRP, UREP, and HSREP grants from QNRF and internal grants from Qatar University, and is involved in MRC grants. He has been involved in EPSRC, ISIF, and EPSRC-ACC grants along with different national and international projects. He has worked as a Consultant for the projects titled ''Driver Distraction Management Using Sensor Data Cloud (2013–2014, Information Society Innovation Fund (ISIF) Asia).'' He received the 2013 ISIF Asia Community Choice Award for a project titled ''Design and Development of Precision Agriculture Information System for Bangladesh.'' He has recently won the ''COVID-19 Dataset Award'' and the ''National AI Competition Awards'' for his contribution to the fight against COVID-19. He is serving as an Associate Editor for IEEE ACCESS and a Topic Editor and a Review Editor for *Frontiers in Neuroscience*.