

# A Novel Hybrid DL Model for Printed Arabic Word Recognition based on GAN

Yazan M. Alwaqfi<sup>1</sup>, Mumtazimah Mohamad<sup>2\*</sup>, Ahmad T. Al-Taani<sup>3</sup>, Nazirah Abd Hamid<sup>4</sup>  
Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia<sup>1, 2, 4</sup>  
Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid, Jordan<sup>3</sup>

**Abstract**—The recognition of printed Arabic words remains an open area for research since Arabic is among the most complex languages. Prior research has shown that few efforts have been made to develop models of accurate Arabic recognition, as most of these models have faced the increasing complexity of the performance and lack of benchmark Arabic datasets. Meanwhile, Deep learning models, such as Convolutional Neural Networks (CNNs), have been shown to be beneficial in reducing the error rate and enhancing accuracy in Arabic character recognition systems. The reliability of these models increases with the depth of layers. Still, the essential condition for more layers is an extensive amount of data. Since CNN generates features by analysing large amounts of data, its performance is directly proportional to the volume of data, as DL models are considered data-hungry algorithms. Nevertheless, this technique suffers from poor generalisation ability and overfitting issues, which affect the Arabic recognition models' accuracy. These issues are due to the limited availability of Arabic databases in terms of accessibility and size, which led to a central problem facing the Arabic language nowadays. Therefore, the Arabic character recognition models still have gaps that need to be bridged. The Deep Learning techniques are also to be improved to increase the accuracy by manipulating the strength of technique in a neural network for handling the lack of datasets and the generalisation ability of the neural network in model building. To solve these problems, this study proposes a hybrid model for Arabic word recognition by adapting a deep convolutional neural network (DCNN) to work as a classifier based on a generative adversarial network (GAN) work as a data augmentation technique to develop a robust hybrid model for improving the accuracy and generalisation ability. Each proposed model is separately evaluated and compared with other state-of-the-art models. These models are tested on the Arabic printed text image dataset (APTII). The proposed hybrid deep learning model shows excellent performance regarding the accuracy, with a score of 99.76% compared to 94.81% for the proposed DCNN model on the APTII dataset. The proposed model indicates highly competitive performance and enhanced accuracy compared to the existing state-of-the-art Arabic printed word recognition models. The results demonstrate that the generalisation of networks and the handling of overfitting have also improved. This study output is comparable to other competitive models and contributes an enhanced Arabic recognition model to the body of knowledge.

**Keywords**—Deep learning; convolutional neural network; generative adversarial network; arabic recognition; image processing

## I. INTRODUCTION

Text recognition is among the fundamental computer science technologies, especially in pattern recognition, computer vision, and image processing. Furthermore, in the field of pattern recognition, text recognition is intended to compete against the human capacity to read written text in terms of speed and accuracy by correlating character codes (e.g., Unicode) with characters images (i.e., graphemes). The recognition process is defined as the process in which characters in images of a text are recognised and detected, which are then converted into data and coded that can be understood by the machine [1].

Arabic is known to be in a complex form of characters. Therefore, developing a novel and accurate Arabic printed word recognition model is still available for study. This is primarily due to the Arabic language characteristics, such as typing from right to left, diacritical marks, cursive characters and overlapping. Moreover, several Arabic letters contain dots, Hamza, Madda, and diacritics, which are markings placed above or below the Arabic letters; any misrecognition of these dots may result in the incorrect representation of the character and, hence, the entire word [2], [3]. Furthermore, multidimensional challenges (language-independent) are faced in developing systems for the Arabic language, which include the lack and absence of Arabic text databases, low-scanning resolution images, font variations, and the complex layout of scanned images. These limitations above contribute to the failure to apply the techniques developed for other types of languages to the Arabic language.

On the other hand, the term "deep learning" refers to a specific kind of machine learning based on an artificial neural network inspired by the structure and function of the human brain [4]. It was established to assist machine learning in achieving one of its primary intentions, which is artificial intelligence. It relies on imitating or mimicking the process in which the human brain processes data and extracting patterns to learn and perform intelligent decisions. Furthermore, deep learning is about learning several levels of representation and abstraction that aid in understanding data like images, sound, and text. Deep learning architectures are based on automated learning from features with no extraction or prior knowledge by using several layers to extract the final and needed information from raw data. Deep learning approaches have widespread use in various fields, including computer sciences, bioinformatics, image and pattern recognition, face and voice

\*Corresponding Author.

recognition, medical image analysis, drug discovery, and health risk evaluation [5]–[8].

Meanwhile, deep learning models such as CNNs have been shown to be beneficial in reducing the error rate and enhancing accuracy in Arabic printed word recognition systems [9]. In contrast, these models have several drawbacks when dealing with Arabic printed word recognition, including poor generalisation ability [2][10], overfitting [11][12] and low reliability [9]. These drawbacks are due to the reliability of Deep learning models, which increases with the depth of layers, but the essential condition for more layers is an extensive amount of data. Since CNN generates features by analysing large amounts of data, its performance is directly proportional to the volume of data, as DL models are considered data-hungry algorithms. The number of trainable parameters in deep learning models runs into millions; thus, appropriate training can only be executed by utilising a large amount of data, meaning that as the number of data increases, the model's performance becomes better behaved. Meanwhile, the lack of training data leads to overfitting, which will generate faulty or inaccurate predictions and decreased reliability. Furthermore, the accuracy rate for Arabic printed word recognition systems stated by the majority of studies using deep learning models has reached very high levels, even though the suggested solutions might not be scaled to other problems of a similar type, where most of these systems were tested on small datasets created privately for a particular task, each with its own evaluation protocols and metrics, making direct comparison and objective benchmarking impractical and not fair. Additionally, some proposed systems prevented other researchers from comparing their findings by making the tested dataset inaccessible to the public. Thus, these comparisons are considered to have low reliability. These issues are due to the limited availability of Arabic databases in terms of accessibility and size, which led to a central problem facing the Arabic language nowadays [13], [14]. Therefore, the availability of large and benchmark datasets is crucial for Arabic printed word recognition.

Overall, the accuracy of Arabic printed word recognition remains a problem in the field of pattern recognition, given that it is still at an infant stage for the Arabic language compared to Latin-based languages. These challenges are attributed to the fact that most Arabic recognition research continues to face the increasing complexity of the system's performance, where few efforts have been made to enhance models of recognising Arabic printed words. This depicts that there is still vast room for enhancements [15] [16].

Using novel deep learning techniques can revolutionise the interest and surge in the field of character recognition. Recently, GANs have been shown to be incredibly effective in various image-processing applications. Concurrently, GANs can be utilised in numerous fields, such as generating high-definition images from low-definition images, producing photo-realistic images of objects, and transforming images across domains [17].

Few studies related to Arabic recognition using GAN have been developed. The adoption of the Convolutional Neural Networks (CNNs) and Generative Adversarial Networks

(GANs) in the proposed research comes from their unintuitive generalisation behaviours and their ability to adapt to the new unseen data that arise from the same distribution as that used when the model was learned. In this paper, we investigate the potential of GANs algorithms in generating text to synthesise a result that could be used to estimate the prospect of the research topic. To build an Arabic printed word recognition model, the proposed methodology focused on enhancing the accuracy and reducing error rates for all possible reasons by covering the most critical phases of Arabic recognition (classification phase and training phase), and this was accomplished by utilising deep learning neural network models (DNN); two GAN models, and one DCNN model.

## II. RELATED WORK

Recent advances in deep learning algorithms, as well as the outstanding results obtained from CNN in image classification and prediction [6], paved the way for researchers to apply them in the classification and recognition of Arabic printed words [18].

In [19], a model is proposed for recognising offline Arabic printed documents to resolve the issue of segmentation for the Arabic text. A pipeline of three neural networks is proposed; the first network model predicted the Arabic words' font size, which is then employed in training the subsequent two models upon normalising the word of 18 points font size. Then words are segmented into characters by the second network model. Meanwhile, the Arabic characters are recognised by CNN by utilising the segmented characters as its input. The significant features are then automatically extracted. To enhance the model's generalisation ability and decrease the overfitting problem, the authors increased the training data size by applying data augmentation techniques. An accuracy rate of 94.3% is achieved when assessed on the APTI dataset.

A deep hybrid learning model is proposed in [20] to recognise printed Arabic text in a variety of font types and fonts that imitate Arabic handwritten scripts. Two bi-directional short long term memory networks (BDLSTM) and five convolution neural networks (CNNs) are employed. The proposed model functioned end-to-end and segmentation free. The proposed model was tested on the APTI database and reached an accuracy of 94.32%.

Recently, research on deep learning for Arabic text recognition has gained more attention, even though available Arabic databases remain limited in accessibility and size. In contrast, researchers have now begun to see artificial data as a safe and practical alternative to overcoming the obstacles to data access. Although generative models have received a lot of attention in the field of machine learning, they had a limited influence before the development of GANs, which provide an alternate data source for the current Arabic recognition systems that use machine learning, allowing the addition of artificial samples to training data to enhance model generalisation. Several GAN variants have been suggested after its advent in various fields, such as image processing and picture super-resolution. Despite this, only a few works based on GANs have been applied to the Arabic language [21]–[24].

According to [24], GANs are a merging technique to learn deep representation without much-annotated training data. The GAN model comprises a pair of deep neural nets aimed at training two adversarial networks by using a form of a min-max game. These adversarial networks include an expert known as a discriminator and a forger known as a generator. Additionally, the generative model receives random noise vectors as input and strives to generate an output (fake images) similar to genuine images (real images). The generator (forger) aims to generate forgeries realistic images that can trick the discriminator.

Meanwhile, the discriminator or expert strives to differentiate between fake and genuine images by classifying forged images generated by the generator as fake and genuine images from the original sample as real. On the other hand, the generator tries to reduce (minimise) its loss by optimising the objective function while the discriminator seeks to increase (maximise) it. A scalar likelihood that the input belongs to real data distribution is what it produces as an output [22], [25]–[27].

GANs are proposed to address the shortcomings of other deep learning algorithms. As mentioned early, the basic principle of GANs is the ability of the generator to generate as realistic samples as possible to deceive the discriminator. In contrast, the discriminator seeks to discern between fake and real examples. Through adversarial learning, the generator and discriminator both get better [27], [28]. Therefore, this adversarial process gives GANs a notable stand out among other generative algorithms. More precisely, GANs are superior to further deep learning algorithms because they can parallelise the generation, which is not achievable with other deep learning techniques like PixelCNN, and the generator design has few restrictions.

On the other hand, GANs are an effective subclass of generative models, which can generate entirely new valid data without requiring mathematical assumptions or the knowledge of explicit true data distribution. These benefits make GANs ideal for numerous astounding implementations that provide unexpected and unseen results; most of these implementations concern image processing [29], [30].

Furthermore, generative modelling is regarded as a strategy for data augmentation, also known as GAN-based data augmentation, and that refers to constructing artificial instances from a dataset that preserves comparable features to the original set. The main objective is to expand the dataset to address the overfitting issue and enhance the generalisation ability [22], [31].

In [22], a SentiGAN model is proposed to address the lack of dialectal Arabic datasets by augmenting low-size datasets and producing a variety of high-quality sentences for five different dialects of Arabic. Five generators and one discriminator are utilised in the dialectal Arabic generating process for a particular dialect, i.e., one generator for each dialect. The generator is used to produce new samples (text), and the discriminator assesses those samples (text), along with a dynamic update, to ensure that the process operates automatically and unsupervised. Before enriching the desired datasets, two metrics; novelty and diversity, are utilised to

check the consistency and the quality of the produced dialectal Arabic text. The MADAR dataset is used, and more sentences are generated than in the original dataset. Experimental results show that the created datasets are reliable and valuable.

The conditional deep convolutional generative adversarial network is the model suggested by the study in [23], it used for the generation of Arabic handwritten isolated characters. The suggested model consists of two CNNs, a discriminator and a generator, and competing to learn. The discriminator is a DCNN architecture composed of four convolutional layers, followed by an FC layer. Each convolutional layer involves batch normalisation and dropout with Leaky ReLU activation. The sequence of convolutions' high-level output is flattened, joined to the one-hot encoded conditional formation, and sent to a dense (FC) layer before being output. The generator is a generative algorithm that attempts to predict each pixel value of an image; the width and height of the input have to be increased to achieve the required shape. The first layer's output is modified and sent through three further layers of transposed/fractionally strode convolutions to predict a single-channelled 32×32 image. Batch normalisation with ReLU activation is performed on each transposed convolutional layer. The proposed model is trained on the AHCD dataset. Experimental results showed the effectiveness of the proposed model. Qualitative and quantitative data demonstrate that the produced samples are relatively equivalent to the genuine ones in terms of diversity and quality. Thus, the proposed model provides an easy substitute for the restricted Arabic handwritten characters database.

GANs have a wide range of types, such as BasicGAN, WGAN, DCGAN, VanillaGAN and BiGAN. These types share a discriminator and generator as standard GAN components, but their architectural designs vary. Consequently, different GANs generate various outputs. These types of GANs are used by [32] to generate different Arabic handwritten characters. The researchers split the proposed approach into two parts, one for generating an image and the other for evaluating the quality of the generated images by assessing how realistic and fake the generated images are. Fréchet inception distance (FID) and native-Arabic human evaluations are utilised. Each GAN is trained separately and generates different outputs due to the differences in their architectures. A total of 40,485 Arabic handwritten characters are generated. According to experimental findings, WGAN performed better in FID, with an accuracy of 96%. In contrast, DCGAN performs better in native-Arabic human evaluation, with an accuracy of 35%.

Furthermore, GANs have shown outstanding results in automatically synthesising high-resolution, realistic images from text representations. In [33], DF-GAN and AraBERT architectures are combined to generate images conditioned with Arabic text descriptions. The initial step is to translate the instruments from English to Arabic on CUB and Oxford-102 flowers datasets and recreate a new dataset that fits utilising the Arabic text-to-image generation operation by using the DeepL-Translator on text descriptions. Secondly, the AraBERT-generated sentence vector dimension was decreased to match the input shape of the FD-GAN. Thirdly, AraBERT is merged with DF-GAN by feeding the sentence embedding

vector to both the DF-GAN generator and discriminator. The proposed approach is evaluated with FID and IS. According to experimental results, the proposed model obtained an accuracy of 3.21% and 3.01% on SI and an accuracy of 60.96% and 65.45% on the FID score in the CUB and Oxford-102 datasets, respectively.

In [24], a novel GAN-based adaptive data augmentation model is proposed to recognise Arabic offline handwritten text. The proposed model is divided into two parts; in the first part, the authors employed an adaptive data augmentation method to generate a balanced dataset. Then, a GAN-based words synthesiser was trained in the new dataset to create images of Arabic words that look like they were written by humans, thereby boosting class diversity. The recurrent neural network library (RNNLIB) is used to develop the experimental models. The proposed model is evaluated on the IFN/ENIT and AHDB datasets and achieved an accuracy of 97.15% and 99.30%, respectively.

Based on the studies mentioned above, only a few of them are conducted on recognising the Arabic language by exploiting GANs to improve the recognition models and enhance accuracy rates, with generally satisfying results. Nevertheless, this is a worthwhile observation about the research in this field. Motivated by these research opportunities, the scarcity of studies and research motivated us to explore the strength of GAN to use it to recognise Arabic printed words to achieve promising results, as this study is considered among the first attempts in this field.

### III. MATERIALS AND METHOD

This section presents the proposed methodology of two deep learning-based models for Arabic printed word recognition; the DCNN model and the hybrid deep learning model based on GAN as a data augmentation technique. Also, it discusses the framework design of the proposed models.

#### A. Dataset

The training datasets are very crucial in the proposed approach since it is a straightforward way to get started with deep learning, and without having a good and sufficient training dataset, the system would never produce results with a high accuracy level. This study utilised the Arabic printed text image (APTI) dataset for testing and training. The APTI dataset is a collection of images of Arabic printed words. It was recently published by [34] for large-scale benchmarking of open-vocabulary, multi-font, multi-size and multi-style text recognition systems in Arabic. There are a total of 113284 different single words in the APTI dataset, each available in 10 diverse fonts: Andalus, Advertising Bold, Arabic Transparent, DecoType Thuluth, Tahoma, Traditional Arabic, Simplified Arabic, M Unicode Sara, and DecoType Naskh. It also comes with 10 different font sizes and four different styles: font sizes (6-18 and 24 points) and styles (plain, italic, bold, and a combination of italic and bold). For the present study, the dataset was split into training, validation and testing sets based on a ratio of 80%, 10%, and 10%, respectively.

#### B. Pre-processing Step

The success of classification for any DNN might be impacted by the pre-processing of the dataset images [35], affecting the final findings' recognition accuracy rate. In this work, both the DCNN model and the hybrid GAN-based model share the same pre-processing step, which involves dataset image resizing and image grey-scaling.

The training dataset images are comprised of various sizes. Therefore, this requires standardising the height and width of these images to cope with a dataset of images that could differ in size (height and width in pixels) throughout the stages of developing DNN layers. The size of the images should adjust before being used as input to the model. Therefore, images were resized to the size of 32×32 pixels.

On the other hand, the issue with images possessing multiple colour channels is that machine learning algorithms, or DNNs, have to function with three different data (R-G-B) values in order to extract the images' features and categorise them. This increases the computational complexity of the operation. DCNNs are responsible for transforming complex images into simpler ones for faster processing without missing pertinent features that are crucial to making an accurate prediction. Consequently, images are transformed into grayscale.

#### C. Proposed DCNN Model

The proposed DCNN model is based on two consequent convolutional layers, each possessing 64 filters, 3×3 kernel size and stride 1. Each layer is followed by a batch normalisation layer and a pooling layer of 2×2. Then the output of these layers is flattened by a flatten layer and squashed into two consequent fully connected dense layers, each possessing 256 neurons. Excluding the last layer, which used the softmax activation function, all other convolutional and dense layers in the proposed DCNN model employed the Leaky ReLU activation function. Lastly, the output of the two dense layers is keyed into the output layer – a fully connected dense layer that uses the softmax activation function to produce the results. The last layer comprises  $n$  neurons that represent the number of output classes. Fig. 1 depicts the overall architecture of the proposed DCNN model. For the optimisation process in the DCNN model, the Adam optimiser is adopted.

Moreover, the sparse categorical cross-entropy function is utilised for a loss function. To achieve the enhancement of the DCNN model, three significant modifications were made to the DCNN architecture. First, the DCCN model is modified by adding a batch normalisation layer after each max pooling layer in the model to catalyse and speed up training while utilising higher learning rates, which simplifies the learning process. Second, the Leaky ReLU activation function is adopted to replace the regular ReLU function in the dense and convolutional layers of the DCNN model to catalyse and faster the training duration, excluding the output layer, which utilises the softmax activation function. The final modification step is applying the sparse Categorical cross-entropy as a loss function to solve the problem of conversion long time.

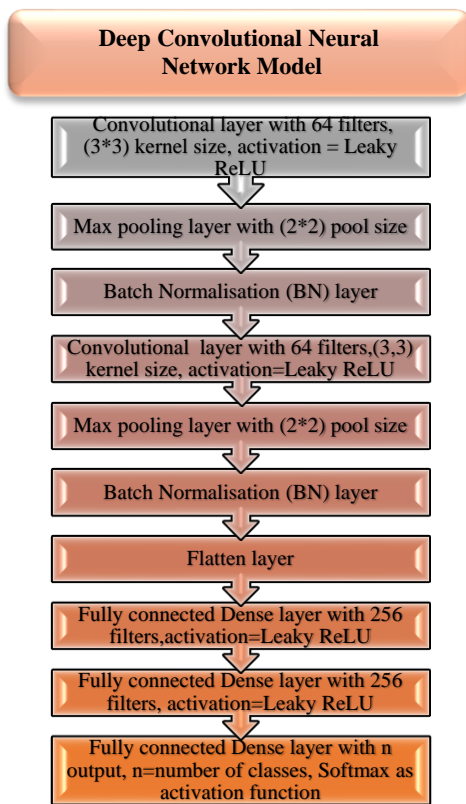


Fig. 1. DCNN architecture

The proposed DCNN model consists of ten layers. Fig. 2 shows the general framework flow of the adopted DCNN model.

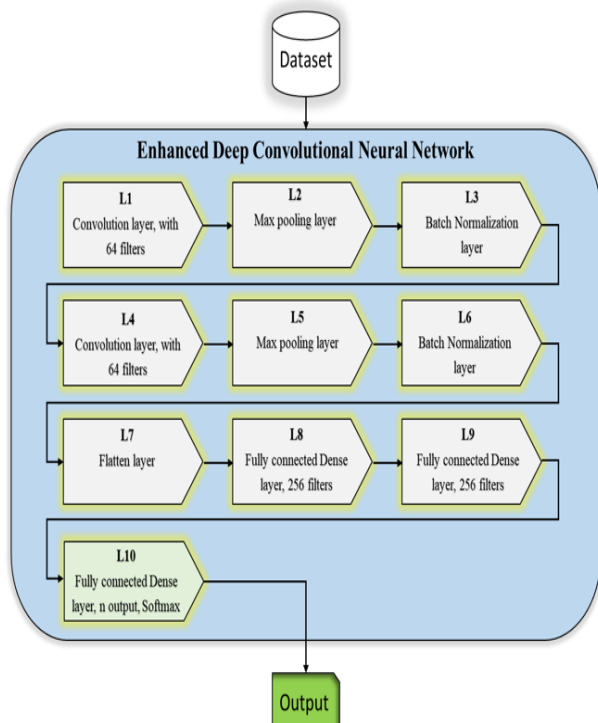


Fig. 2. DCNN model framework flow

As shown in Fig. 2, the proposed DCNN model layers are listed in Table I.

TABLE I. DCNN MODEL LAYERS

Layer no.	Layer Description
i	First layer is a 2D convolution layer (Conv2D). It uses 64 filters with a 3x3 kernel size. The training phase is catalysed and sped up by utilizing the Leaky ReLU as the activation function while preventing consistent predictions for negative input values.
ii	The second layer is a max polling layer with a 2x2 pool size. A pooling layer is a new layer added after the convolutional layer, particularly after nonlinearity, for instance, following applying Relu on the feature maps output by a convolutional layer. The dimensions of the feature maps can be reduced by using the pooling layers. Hence, the number of parameters required to learn and the amount of computation performed in the network are reduced.
iii	The third layer is the Batch Normalization layer. In each iteration, it normalizes the inputs by subtracting their mean and dividing by their standard deviation, and then it applies a scaling coefficient and a scale offset. After applying standardization, the resulting mini-batch has zero mean and unit variance.
iv	A 2D convolution layer (Conv2D) comprising a 3x3 kernel size and 64 filters makes up the Fourth layer. The layer's training was sped up by applying the Leaky ReLU as an activation function.
v	The fifth Layer is another max polling layer with a 2x2 pool size.
vi	The sixth layer is another Batch Normalisation layer.
vii	The seventh Layer is a Flatten layer, which is utilized to map the output of the previous Conv2D layer. After that, the Conv2D layer is flattened to be applied in the Dense layer as input. Some Neural Network implementations might not be able to map a spatial structure directly into a dense layer; thus, a flatten layer is necessary in between. Therefore, if the rank is higher, like with convolutions, the layer is implicitly flattened at first. So, the Flatten layer moves from convolution 2D to the Dense layer.
viii	The eighth layer is a fully connected dense layer using 256 filers and implementing Leaky Relu as an activation function.
ix	The ninth layer is a fully connected dense layer using 256 filers and implementing Leaky Relu as an activation function, which comprises a fully connected layer to transition from feature maps to an output prediction for the model.
x	The tenth layer is the output layer for the enhanced DCNN model. This layer uses the dense layer with the number of N+1 classes, and it implements the Softmax as an activation function.

Overfitting is one of the core difficulties that occur in multilayer NN learning, especially in deep neural networks. Overfitting generally happens once the DNN model performs effectively on the training dataset but fares badly on the testing dataset. Batch normalisation and dropout are two well-known approaches to overcoming this obstacle. Despite the fact that the two approaches share overlapping design fundamentals, multiple study findings have proven that they each offer particular advantages for enhancing deep learning.

However, several recent works have shown that the dropout technique is ineffective or even detrimental to CNNs training, which has proven that there are drawbacks to use in convolutional layers and sometimes its inefficiency. In

contrast, the training duration is typically increased upon applying dropout to a neural network, given that dropout increases the size of the network, which eliminates units during the training phase, and reduces the network capacity [11], [12].

On the other hand, adding batch normalisation in DNN architecture significantly improves the accuracy. Furthermore, batch normalisation substantially reduces training time and permits the utilisation of high learning rates. As a result, the training steps that are needed for network convergence are reduced [36].

According to [37], adding batch normalisation layers after the non-linear layers improves accuracy. Before introducing batch normalisation (BN), the network's training time was substantially dependent on the precise hyperparameters initialisation, including the adoption of modest learning rates, thereby increasing the time of training. During DNN training, each layer's input value distribution is influenced by the layers which came before it. Because this variation slows training, BN was developed to address it and accelerate the process of learning. The process of BN involves adjusting the unit value of every batch, and since batches are generated at random throughout training, additional noise is introduced into the training process. The noise works as a regulariser, and its effects are like that generated by dropout. Resultantly, dropout can be removed from the deep neural network.

According to [36], the addition of batch normalisation layers should be among the initial steps in optimising a CNN, not just for handling overfitting but also for its ability to improve accuracy significantly. To overcome the overfitting problem in the proposed enhanced DCNN Model, the batch normalisation layers are adopted and used after each convolutional layer in the network model.

#### D. Hybrid Model Based on GANs

This section presents the procedures applied to develop the proposed hybrid GAN-based model. The proposed hybrid GAN-based model combines the generative adversarial networks and the DCNN. The architecture of the proposed hybrid model consists of a GAN model that is used as a data augmentation technique and a DCNN model as a classifier employed for printed word recognition.

First, the GAN model is built, which comprises two deep neural networks; the discriminator network and the generator network. The generator generates new images, whereas the discriminator determines the originality of the generated images. The discriminator evaluates each produced image to decide if it is part of the training set or not. Nevertheless, the generator attempts to generate an additional fake instance in order to convince the discriminator that the image generated corresponds to the training set. During the training process, the discriminator and generator models compete with one another. While the generator strives to fool the discriminator, the discriminator attempts hard to avoid being deceived.

As a result, this competition between the two models encourages both to enhance their capabilities. The skills of these neural networks are sharpened as they undergo more training, thus increasing the ability of the generator to generate

real-like images which indistinguishable from that observed in the real world. Concurrently, the discriminator is more effective in detecting fake images. The accuracy of the final results will improve as the size of the training dataset increases. The training for the dataset will align with the level of the Arabic printed words. Fig. 3 shows the overall research framework for the proposed hybrid GAN-based model.

The second step entailed the classifier DCNN model development. The proposed DCNN is utilised to classify the images of Arabic printed words, which is the last step in the proposed hybrid GAN-based mode. The DCNN is employed, given that it is a typical multiclass classification NN. The classifier DCNN accepts the new dataset as input and captures various perspectives on features via the convolutional operation. The resulting feature maps produced from each convolution layer are passed to the following convolution layers with extra kernels. The reason for the pass is to extract a higher level of features of the input image. The pooling layer then summarises the feature map features generated by the convolution layers to reduce the dimensionality of the feature maps, hence minimising the amount of computation performed in the network. The classifier DCNN in the proposed hybrid GAN-based model shares the same architecture as the proposed DCNN model introduced previously. Fig. 1 shows the shared architecture used for both the stand-alone enhanced DCNN model and the classifier DCNN in the proposed hybrid GAN-based model. The DCNN model classifier uses the newly generated collection of images. This way, generative adversarial modelling is used as an alternative to data augmentation techniques in the proposed hybrid GAN-based model.

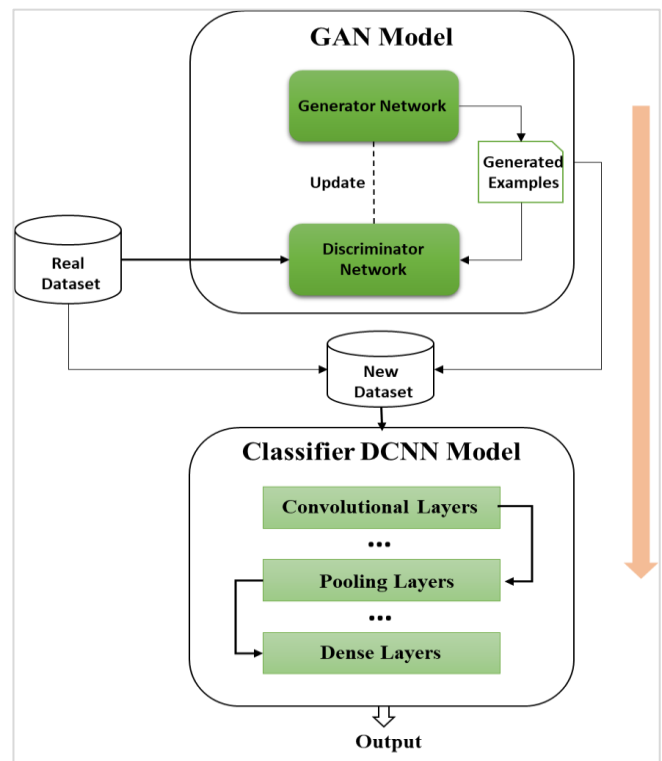


Fig. 3. Hybrid GAN-Based model design

1) *Generator neural network:* The proposed hybrid GAN-based model employs a generator neural network with ten layers. Two of the proposed enhancements for the proposed DCNN model are adopted in the generator neural network, the first one is the addition of adding batch normalisation layers, while the second one is adopting the Leaky ReLU as an activation function. The generator neural network layers are listed in Table II.

TABLE II. GENERATOR NEURAL NETWORK LAYERS

Layer no.	Layer Description
i	A sequential neural network of 10 layers is built with 8 layers as the hidden layers.
ii	The first layer is the input layer – a fully-connected dense layer with 32*32*256 filters, 32*32 for the width and height and 256 neurons.
iii	Second layer is a batch normalisation layer.
iv	Third layer is a reshape layer to convert the shape of data into two-dimensional representation.
v	Fourth layer is a transposed convolution layer with 256 filters and (3*3) kernel size.
vi	Fifth layer is another batch normalisation layer.
vii	Sixth layer is transposed convolution layer with 128 filters and (3*3) kernel size.
viii	Seventh layer is another batch normalisation layer.
ix	Eighth layer is another transposed convolution layer with 64 filters and (3*3) kernel size.
x	Ninth is the last batch normalisation layer in the network.
xi	The last layer is the output layer, which is another transposed convolution layer with (3*3) kernel size.
xii	The generator neural network is compiled using Adam Optimizer.
xiii	The loss in the generator neural network can be found with cross-entropy loss function.

2) *Discriminator neural network:* The discriminator neural network for the proposed hybrid GAN-based model consists of nine layers. For the additional enhancements, the batch normalisation layers are added after each convolution layer, and Leaky ReLU is adopted as an activation function. The discriminator neural network layers are listed in Table III.

TABLE III. DISCRIMINATOR NEURAL NETWORK LAYERS

Layer no.	Layer Description
i	A sequential neural network of 9-layers is built with 7 layers as the hidden layers.
ii	First layer is a convolution layer that has 64 filters and 3*3 kernel size with input shape 32*32*1 for height, width and one-color channel.
iii	Second layer is another convolution layer that has 128 filters and 3*3 kernel size
iv	Third layer is layer is another batch normalisation layer.
v	Fourth layer is another convolution layer with 256 filters and (3*3) kernel size.
vi	Fifth layer is another batch normalisation layer.
vii	Sixth layer is known as the flatten layer, this layer used to resize and flatten the data.
viii	Seventh layer is fully connected dense layer with 128 neurons (filters).
ix	Eighth layer (the last hidden layer) is another fully connected dense layer with 128 neurons (filters).
x	Ninth is the last batch normalisation layer in the

	network.
xi	The final layer is the output layer, a fully-connected dense layer with n+1 neurons (n = the number of classes).
xii	The discriminator neural network is compiled using Adam Optimiser.
xiii	The loss in the discriminator NN can be found with cross-entropy loss function.

3) *Deep CNN classifier:* The DCNN classifier in the proposed hybrid GAN-based model has the same architecture as the proposed DCNN model that introduced in section C. As mentioned previously, the DCNN classifier for the proposed hybrid GAN-based model consists of ten layers in total, and it has the same additional enhancements as the proposed DCNN model. The additional enhancements were the batch normalisation Layers, adopting the Leaky ReLU as the activation function and using sparse categorical cross-entropy as a loss function.

4) *Handling Overfitting for The Hybrid GAN-Based Model:* Since the proposed model consists of two parts; the GAN and the classifier DCNN part, two techniques were adopted to handle the possibility of overfitting in any of the two parts:

a) *Handling overfitting for the Generative Adversarial Part:* To avoid overfitting, generative and discriminative algorithms are used. The training dataset is fed to the discriminator; hence, the generator lacks any information on the training dataset while it creates new data instances. The generator is fed only from the discriminator, which decides whether each instance of data reviewed belongs to the actual training dataset or not. The proposed model is designed in a way that the generator never sees the genuine data; it must learn to create realistic information by receiving feedback from the discriminator. This process is called adversarial loss and works surprisingly well when implemented correctly.

b) *Handling overfitting for the Classifier Deep Convolutional Neural Network Part:* The significance of the proposed models in using GANs is to produce more realistic images to increase the dataset size for training the classifier deep CNN network. Among the techniques to handle overfitting is having more training datasets. Algorithms can be supported to detect signals effectively by training with more datasets. However, the technique would not be helpful if just noisier data were added to the training data. Thus, it should ensure that the newly generated dataset of images is clean and relevant in the GAN model. Actually, this technique works, such as data augmentation techniques, which can apply to increase the dataset's size artificially. Data augmentation refers to increasing the data size, which is increasing the number of images available in the dataset. As a result, in this way, generative adversarial modelling is utilised as an alternative to data augmentation techniques in the proposed hybrid GAN-based model. The use of generative adversarial modelling as a data augmentation technique offers a more domain-specific way to increase the number of training images with new high-quality generated images in domains with limited data. This approach enables the model to address the overfitting problem.

In addition to the previous solution, batch normalisation layers are adopted and added to the classifier DCCN part to tackle the overfitting problem in the proposed model. Since the DCNN classifier shares the same architecture as the DCNN model that proposed, a batch normalisation layer was added after each convolutional layer in the classifier DCNN to make an additional step in overfitting handling for the overall proposed hybrid GAN-based model.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

Two models have been conducted in this study. The first model conducted is the DCNN model. Then the second model is a hybrid deep learning model based on GAN. In addition, this experimental design discusses the aim of the experiments, as well as how they may be attained. The two proposed models are evaluated using the Arabic printed text image database (APTI). The dataset is split into three sets (training, validation, and testing) using a ratio of 80%, 10%, and 10%, respectively. The implementation part of the proposed approach is a program written in Python using TensorFlow 2.0 and Keras.

##### A. DCNN Model

To evaluate the accuracy of the proposed DCNN model and to compare it with other approaches, the model is analysed through a number of evaluation scenarios using APTI datasets. The proposed DCNN is tested on the dataset before and after the additional enhancements. The results obtained from each experiment for the proposed DCNN model were compared before and after the enhancements.

Regarding the testing results, the proposed enhanced DCNN model achieved high (WRR) accuracy scores on the APTI dataset, which corresponded to an average overall test set and validation accuracy of 94.81% and 94.96%, respectively. On the other hand, the DCNN model without enhancements achieved an average accuracy of 93.67% and 94.17% for validation. The differences in results are shown in Fig. 5. The enhanced model on the APTI dataset starts learning with fast steps from the first epoch to the 20<sup>th</sup> epoch in a continuous increase in the accuracy until it attained an accuracy of 91%. Then, the training process started to slow down after the 20<sup>th</sup> epoch to the 40<sup>th</sup>. After that, the model reached stability in the training accuracy at epoch number 40. The model shows the accuracy results in ascending values and achieves the best accuracy in the 40 epochs with training accuracy.

From Fig. 4, the enhanced DCNN model learns better than the original DCNN model in the training phase. CNN models learn better if the training process is better; that means the enhancements of adding batch normalisation Layers, adopting the leaky ReLU as an activation function, and using sparse categorical cross-entropy as a loss function helped to make the model training process faster and more accurate in the Arabic printed words recognition and classification phase. Furthermore, the inclusion of batch normalisation layers facilitates the usage of higher learning rates during the training process. This explains why the enhanced DCNN model performed better in testing and validation results of the

enhanced DCNN model than the standard DCNN model without enhancements.

On the other hand, the proposed enhanced DCNN model for Arabic printed word recognition achieved high recognition accuracy (WRR) of 94.81%, outperforming other state-of-art models. Compared to the recognition accuracy obtained by the hybrid deep learning model based on CNN and BDLSTM to recognise printed Arabic text proposed by [15], where the enhanced DCNN model outperformed the hybrid CNN and BDLSTM Model by a small margin and achieved a recognition accuracy of 94.32%. In addition, the work by [14] proposed three neural networks pipeline model to recognise printed text images of the APTI dataset of words that contains over two-million-word samples, which uses data augmentation techniques and was able to score a recognition accuracy of 94.30%. The detailed comparison between the two-word recognition models is illustrated in Fig. 5.

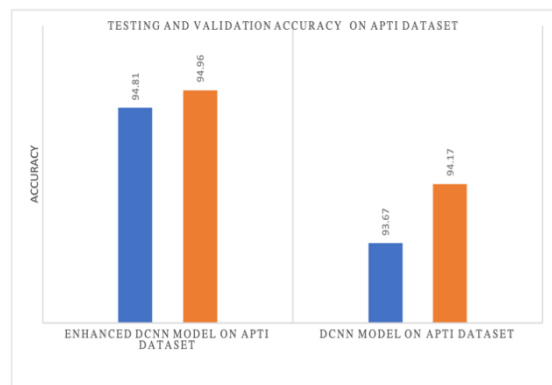


Fig. 4. Testing and Validation Accuracy for the DCNN Model with and without the enhancements

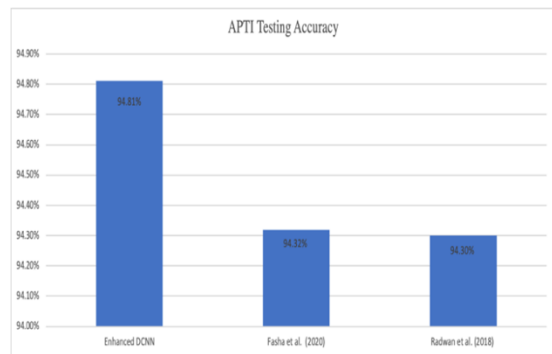


Fig. 5. Testing Results for the Enhanced DCNN against State-of-Art Models

##### B. Hybrid DL Model based On GAN

The performance of the proposed hybrid GAN-based model is evaluated by introducing the GAN phase into the proposed enhanced DCNN model to generate a new dataset of clean and relevant images. This occurs using GAN as a data Augmentation technique by artificially increasing the dataset size. The experimental testing for the proposed hybrid GAN-Based model is conducted in APTI datasets. Firstly, the results obtained from the proposed hybrid GAN-Based model were compared to those from the enhanced DCNN model to determine the impact of using the GAN phase, then compared



the proposed hybrid GAN-Based model with the state-of-the-art studies.

In terms of the testing results, very high accuracy scores (WRR) are achieved by the full proposed hybrid GAN-Based model with generative adversarial step on the APTI dataset. The overall test set accuracy is 99.76%, and 99.85% is achieved for the validation accuracy, while the corresponding values for the proposed enhanced DCNN model without the generative adversarial step as data augmentation technique was 94.81% and 94.96%, respectively. The discrepancies in results are presented in Fig. 6.

Fig. 6 illustrates that the proposed hybrid GAN-Based model achieved a higher success rate on the APTI dataset with higher testing accuracy results compared to the enhanced DCNN model. Thus, the hybrid GAN-based recognition model learnt better than the Enhanced DCNN model in the training phase.

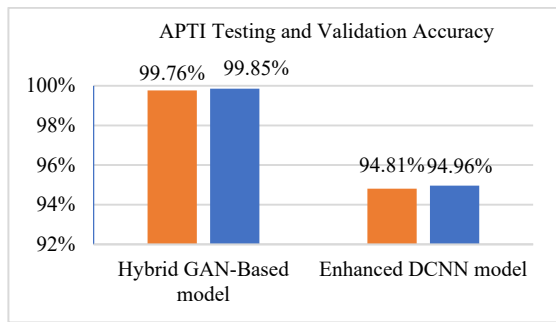


Fig. 6. Testing and Validation Accuracy for the proposed Hybrid GAN-Based model with the Enhanced DCNN model

Furthermore, as depicted in Fig. 7, the proposed hybrid GAN-Based model also achieved promising validation scores over the enhanced DCNN model. The proposed hybrid model that based on using the generative adversarial networks as a data augmentation technique achieved higher validation accuracy than the enhanced DCNN model. The overall validation and testing results of the hybrid GAN-Based model and the enhanced DCNN model on the APTI dataset are presented in Table IV below.

TABLE IV. APTI TESTING AND VALIDATION ACCURACY RESULTS

Implementation	Validation Accuracy %	Test Accuracy %
Hybrid GAN-Based model	99.85	99.76
Enhanced DCNN model	94.96	94.81

From the previous Fig. 6 and Table IV, it is notable that the hybrid GAN-based model that is based on using the GANs as a data augmentation technique outperformed the regular enhanced DCNN model in the APTI dataset, producing higher validation and test accuracy results. Adding GANs to the classifier DCNN model and using it as a data augmentation technique assisted in making the model training process more accurate in Arabic printed word recognition and classification.

As observed in the experiments, the hybrid GAN-based model commenced learning with fast steps from the first epoch to the 20<sup>th</sup> epoch with continuously increasing accuracy until a

value of approximately 93% is attained. Then, the training process started to slow down after the 20<sup>th</sup> epoch to the 40<sup>th</sup> one. After that, the model reached stability in the training accuracy at epoch number 40. The hybrid GAN-based model shows the result of accuracy in ascending values, achieving the best accuracy in the 40 epochs with training accuracy. The testing and validation accuracy measures utilise the word recognition rate (WRR) as a measurement unit on the APTI dataset.

On the other hand, the hybrid GAN-based model for Arabic printed word recognition achieved high recognition accuracy of 99.76%, thereby outperforming other the best state-of-the-art models. The proposed hybrid GAN-based model outperformed the recognition accuracy values obtained by the hybrid CNN and BDLSTM model proposed by [20] on 18 fonts and duplicate words across fonts types in the APTI dataset with a very good margin (99.76% vs. 94.32%), as well as the three NN pipeline model proposed by [19], that used traditional data augmentation techniques regarding recognition accuracy on the Printed Text Images of the APTI dataset and achieved recognition accuracy of 94.30%. The detailed comparison between the printed word recognition models is illustrated in Fig. 7.

The experimental results for the proposed hybrid GAN-based model demonstrated that the proposed model achieved very high accuracy (WRR) scores on the APTI dataset. Specifically, the proposed hybrid GAN-based model achieved average overall testing and validation accuracy of 99.76% and 99.85%, respectively. Meanwhile, the proposed enhanced DCNN model, without using any additional data augmentation techniques, achieved an average overall test set accuracy of 94.81% and 94.96% for the validation accuracy. The proposed hybrid GAN-based benefits from all the additional enhancements that are added to the enhanced DCNN model, batch normalisation layers were added to the proposed hybrid GAN-based model networks after each convolutional layer. The leaky ReLU was also utilised as an activation function, and the sparse categorical cross-entropy was employed in the classifier DCNN part of the proposed hybrid GAN-based model as the loss function.

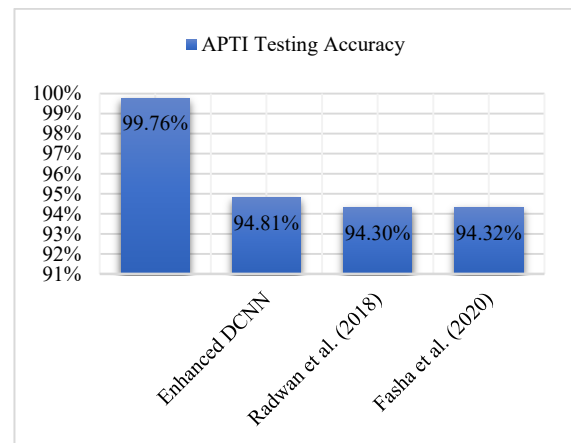


Fig. 7. APTI Testing Results of the proposed Hybrid GAN-Based model and other state-of-the-art models

From an analytical point of view, Fig. 6 and Fig. 7 depict the highly competitive testing accuracy results achieved by the proposed hybrid GAN-Based model, which outperformed other state-of-the-art models. The result proves that the proposed hybrid GAN-Based model could learn more and better during the training phase than the other machine learning-based models. The bigger training data generated from the generative adversarial model assisted the proposed hybrid GAN-based model in scoring higher recognition accuracy. This explains why the proposed hybrid GAN-based model outperformed other CNN models in the state-of-the-art models for Arabic printed word recognition on the APTI dataset.

## V. CONCLUSION

This study is considered one of the first to successfully develop a novel framework model based on DL techniques using DCNN and GAN to recognise Arabic words. The proposed model included two kinds of deep neural networks; GAN and DCNN, to enhance the training and classification phase of the printed word recognition model. The generative adversarial neural networks were designed to work as a data augmentation technique that is applied to artificially increase the size of the dataset by generating a new set of images and adding it to the original dataset to tackle the lack of Arabic datasets, the more extensive dataset will enhance the training accuracy for the Arabic text recognition classifier DCNN model. Additionally, solutions for the overfitting problem were introduced, such as having more data to train by employing the GAN as a data augmentation technique. More data training can assist models in avoiding overfitting issues. The experimental results showed the strength of the proposed approach of using the GAN model as a data augmentation technique.

The impressive testing accuracy results for printed word recognition prove that generative adversarial models can be used as a very good alternative data augmentation technique that can be applied to artificially increase the size of the dataset by new high-quality and real-like images indistinguishable from the real dataset images. Such new images can be crucial in boosting the training accuracy for the DCNN models. The results of the proposed model are promising, where the accuracy of the proposed model was 99.76%. These results prove the potential of the proposed model to develop an Arabic language recognition field. In future, we will test the proposed model on several languages to evaluate the proposed hybrid GAN-based model in terms of generalisation and token independency.

## REFERENCES

- [1] A. Chhabra, "Deep Learning Based Real Time Devanagari Character Recognition," 2019.
- [2] G. et al. Ahmed, R., "Novel Deep Convolutional Neural Network-Based Contextual Recognition of Arabic Handwritten Scripts," *Entropy*, vol. 23, no. 3, pp. 4–6, 2021, doi: 10.3390/e23030340.
- [3] R. Maalej and M. Kherallah, "Convolutional Neural Network and BLSTM for Offline Arabic Handwriting Recognition," *ACIT 2018 - 19th International Arab Conference on Information Technology*, pp. 1–6, 2018, doi: 10.1109/ACIT.2018.8672667.
- [4] A. Labach, H. Salehinejad, and S. Valaee, "Survey of Dropout Methods for Deep Neural Networks," 2019, [Online]. Available: <http://arxiv.org/abs/1904.13310>
- [5] H. Q. Ghadhbhan and M. Othman, "Survey of Offline Arabic Handwriting Word Recognition Survey of Offline Arabic Handwriting Word Recognition," no. February, 2020, doi: 10.1007/978-3-030-36056-6.
- [6] M. Makhtar, R. Rosly, M. K. Awang, M. Mohamad, and A. H. Zakaria, "A multi-classifier method based deep learning approach for breast cancer," *International Journal of Engineering Trends and Technology*, no. 1, pp. 102–107, Aug. 2020, doi: 10.14445/22315381/CAT13P217.
- [7] N. Altwajry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," *Neural Comput Appl*, vol. 33, no. 7, pp. 2249–2261, 2020, doi: 10.1007/s00521-020-05070-8.
- [8] Oueslati, "A Review of Sentiment Analysis Research in Arabic Language," *Future Generation Computer Systems* 112 (2020): 408-430, 2020.
- [9] L. F. Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al Dujaili, Ye Duan, Omran Al Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al Amidie, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.
- [10] M. Jain, "Unconstrained Arabic & Urdu Text Recognition using Deep CNN-RNN Hybrid Networks," no. July, 2018.
- [11] Cai, Shaofeng, et al., "Efficient and Effective Dropout for Deep Convolutional Neural Networks," *arXiv preprint arXiv:1904.03392*, pp. 1–12, 2019.
- [12] Zeng, Yuyuan, et al. "Correlation-based structural dropout for convolutional neural networks," *Pattern Recognit*, vol. 120, p. 108117, 2021, doi: 10.1016/j.patcog.2021.108117.
- [13] T. Meriem, M. Asma, B. Rahima, K. Said, and B. Belgacem, "Recognition of Printed Arabic Characters".
- [14] I. Saleh Al-Sheikh, M. Mohd, and L. Warlina, "A Review of Arabic Text Recognition Dataset," *Asia-Pacific Journal of Information Technology and Multimedia*, vol. 09, no. 01, pp. 69–81, 2020, doi: 10.17576/apjtitm-2020-0901-06.
- [15] Darwish et al., *Arabic Offline Character Recognition Model Using Non-dominated Rank Sorting Genetic Algorithm*, vol. 1261 AISC, no. October. Springer International Publishing, 2021. doi: 10.1007/978-3-030-58669-0\_20.
- [16] R. Mohd, M., Qamar, F., Al-Sheikh, I., & Salah, "Quranic optical text recognition using deep learning models," *IEEE Access*, vol. 9, pp. 38318–38330, 2021, doi: 10.1109/ACCESS.2021.3064019.
- [17] Creswell, "Generative Adversarial Networks : An Overview," no. April, pp. 1–14, 2018.
- [18] A. Bhardwaj, "An Accurate Deep-Learning Model for Handwritten Devanagari Character Recognition," vol. 7, no. 2, pp. 1317–1328, 2022.
- [19] M. A. Radwan, M. I. Khalil, and H. M. Abbas, "Neural Networks Pipeline for Offline Machine Printed Arabic OCR," *Neural Process Lett*, vol. 48, no. 2, pp. 769–787, 2018, doi: 10.1007/s11063-017-9727-y.
- [20] J. A. Mohammad Fasha, Bassam Hammo, Nadim Obeid, "A hybrid deep learning model for arabic text recognition," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, pp. 122–130, 2020, doi: 10.14569/IJACSA.2020.0110816.
- [21] Y. M. Alwaqfi, M. Mohamad, and A. T. Al-Taani, "Generative Adversarial Network for an Improved Arabic Handwritten Characters Recognition," *International Journal of Advances in Soft Computing and its Applications*, vol. 14, no. 1, 2022, doi: 10.15849/IJASCA.220328.12.
- [22] X. A. Carrasco, A. Elnagar, and M. Lataifeh, "A Generative Adversarial Network for Data Augmentation: The Case of Arabic Regional Dialects," *Procedia CIRP*, vol. 189, pp. 92–99, 2021, doi: 10.1016/j.procs.2021.05.072.
- [23] S. M. S. Ismail B. Mustapha, Shafaatunnur Hasan, Hatem Nabus, "Conditional Deep Convolutional Generative Adversarial Networks for Isolated Handwritten Arabic Character Generation," *Arab J Sci Eng*, no. 0123456789, 2021, doi: 10.1007/s13369-021-05796-0.

- [24] Z. Eltay, et al. "Generative adversarial network based adaptive data augmentation for handwritten Arabic text recognition," *PeerJ Comput Sci*, vol. 8, pp. 1–22, Jan. 2022, doi: 10.7717/PEERJ-CS.861.
- [25] S. Barua, S. M. Erfani, and J. Bailey, "FCC-GAN: A Fully Connected and Convolutional Net Architecture for GANs," no. ii, 2019, [Online]. Available: <http://arxiv.org/abs/1905.02417>
- [26] A. P. Behera, S. Godage, S. Verma, and M. Kumar, "Regularized Deep Convolutional Generative Adversarial Network," *Communications in Computer and Information Science*, vol. 1378 CCIS, no. September, pp. 452–464, 2021, doi: 10.1007/978-981-16-1103-2\_38.
- [27] S. Khamekhem Jemni, M. A. Souibgui, Y. Kessentini, and A. Fornés, "Enhance to read better: A Multi-Task Adversarial Network for Handwritten Document Image Enhancement," *Pattern Recognit*, vol. 123, 2022, doi: 10.1016/j.patcog.2021.108370.
- [28] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications," *IEEE Trans Knowl Data Eng*, vol. 14, no. 8, pp. 1–28, 2021, doi: 10.1109/TKDE.2021.3130191.
- [29] T. Iqbal and S. Qureshi, "The Survey: Text Generation Models in Deep Learning.," *Journal of King Saud University - Computer and Information Sciences*, 2020, doi: 10.1016/j.jksuci.2020.04.001.
- [30] L. Jose, S. Liu, C. Russo, A. Nadort, and A. di Ieva, "Generative Adversarial Networks in Digital Pathology and Histopathological Image Processing : A Review," *J Pathol Inform*, vol. 12, no. 1, p. 43, 2021, doi: 10.4103/jpi.jpi.
- [31] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J Big Data*, 2019, doi: 10.1186/s40537-019-0197-0.
- [32] T. Alkhodidi et al., "GEAC: Generating and Evaluating Handwritten Arabic Characters Using Generative Adversarial Networks," *Proceedings of 2nd IEEE International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2021*, pp. 228–233, 2021, doi: 10.1109/ICCIKE51210.2021.9410746.
- [33] M. Bahani and K. Maalmi, "Fusion of AraBERT and DF-GAN for Arabic Text to Image Generation Fusion of AraBERT and DF-GAN for Arabic," pp. 0–16, 2022.
- [34] F. Slimane, R. Ingold, S. Kanoun, A. M. Alimi, and J. Hennebert, "A new Arabic printed text image database and evaluation protocols," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2009*, pp. 946–950. doi: 10.1109/ICDAR.2009.155.
- [35] Y. M. Alwaqfi and M. Mohamad, "A review of Arabic optical character recognition techniques & performance," *International Journal of Engineering Trends and Technology*, no. 1, pp. 44–51, 2020, doi: 10.14445/22315381/CATI1P208.
- [36] C. Garbin, X. Zhu, and O. Marques, "Dropout vs . batch normalization : an empirical study of their impact to deep learning," pp. 12777–12815, 2020.
- [37] C. S. Sergey Ioffe, "Australian Literary Journalism and 'Missing Voices': How Helen Garner finally resolves this recurring ethical tension," *Journalism Practice*, vol. 10, no. 6, pp. 730–743, 2015, doi: 10.1080/17512786.2015.1058180.