Pushing the Boundaries of Pharmaceutical Drug Classification with Advanced Deep Learning Techniques

Kuppusamy P
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
drpkscse@gmail.com

Yogeeswar Reddy P
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
yogeswar58@gmail.com

Veerendra B
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
veerendra.367874@gmail.com

Sri Satya Aarthi N
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
.com

Pooja S
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
Pooju1672@gmail.com

Abstract— **Capsules and tablets are best and feasible things to store medicines, since people are directly taking those in, so quality of the capsule would maters a lot for our health. By considering the demands of tablet production, this research gives you a drug detection model on the basis of convolution neural network (CNN) which is one major used algorithms in deep learning. Within these yearly years Deep Learning computing paradigm was considered as the best level among the Machine Learning (ML) section. Furthermore, in the ML field it is the most extensively used computational method and got excellent outcomes on various large percipient tasks that match or exceed human performance. The major advantage of the DL is its nature of learn on large and complex data [1]. The DL got used rapidly and large amount in various applications in these years, moreover it got huge success in all the applications those implemented it. Especially in Robotics, cybersecurity, natural language processing, bioinformation and medical information processing etc. Many research works got submitted to review the Avant-grade of deep learning, but every work deal with one side of deep learning and not having a strong base on DL. Therefore, this post proposes to use a more comprehensive approach to provide a better starting point for fully understanding DL. In particular, this abstract try to give a better complete overview of almost every crucial features of deep learning, which has recent improvements included to it. Specifically, this research reviews meaning of deep learning and introduces its methods and network types. Next, this work introduces the most commonly used DL network type, convolutional neural networks (CNNs), and discuss the evolution of CNN architectures and their key features. CNN algorithms help to identify the type and name of the capsules. Images of different types of capsules are gathered for training the CNN models to detect and predict the name and information of capsule. Adam optimizer (Densenet) algorithm launched to train the model fast and to improve convergence of the model. And for measuring the predictive power of the model, cross-entropy has used in the place of loss function for this work. Finally, based on the best parameter for the dense model by cross checking the results with various parameters, the detection method would choose. The results show accuracy of almost 85% above for all the models which are implemented in this research work. Therefore, these methods can be used for automating the process of identification and classification of the capsules. At the end this work concludes with a confusion matrix, benchmark data sets, and a summary and conclusion.**

Keywords—deep learning, convolution neural network, dense network, mobilenetv2, adam optimizer, convergence, cross-entropy.

I. Introduction

Medicine quality is significant in the medical field, and privately tied to people's health. The capsule form of drug and healthcare particulars is pivotal to people's lives. The mortal factors prostration and ignorance are to condemn for the maturity of pharmaceutical crimes. Still, numerous people do not have knowledge to identify the tablet or medicine in terms of disease. It is a pivotal that they have at least minimum understanding of the drug which are required. Look Alike Sound Alike (LASA) is another significant issue in the aspect of druggists or croakers. To help LASA, it's a good idea to modify the names and packaging of the medicines. To discover problematic drug name label combinations, experimenters employed map reviews, and appropriate methods. They also erected an automated discovery system to solve these issues. However, medicine identification is the significant issue due to numerous medicines and tablets appear with similar patterns, relatively complex, and numerous substances need to be linked. The current drugs and tablet packing pattern contains significant downsides in terms of understanding to common people. The medicine industries have installed some assistive tools that bear some pre-processing methods. However, the ministry and technology use the barcode pattern on the medicine to identify the label. In this case, pre-processing was completed by categorizing the colourful capsules and medicines using barcodes. However, the pattern structure is not helping to the public to identify the drugs and tablets. Hence, the researchers are focusing on this domain to help the public in recognizing the appropriate drugs for their illness which are issued by pharmacist. This study focuses on the issue of visual fester packaging plates being utilized to identify medicines. The deep learning-based drug identification model is proposed to help individualities in recognizing the lozenge or specifics, as well as to help druggists with capsule development or generation. The deep learning networks are applied to learn the complications in the problem like visual recognition system which can capture and identify the medicine and capsules. Multiple connection of layers is designed to learn high, medium and low level complex features of given input. The output layer contains the neurons to represent the number of classes for multiclass, and only two neurons for binary class. The standard backpropagation algorithms utilized for decrease the cost, and activation functions such as sigmoid, ReLU, and softmax. Image classification is the process of assigning the appropriate label to the corresponding image using deep neural network.

II. Related Work

Solid oral lozenge represents the pharmaceutical medications which includes the tablets and gelatine capsules. These types of lozenges dominate the business because they deliver a constant controlled lozenge of active medicine. The shape, size and colour are visual characteristics used to recognize and identify the individual products based on illness, and brand Colour tablets are used to help druggists in lozenge administrations. The anticipation of the flavour and colour of a product is important for illustration, consumers anticipate a cherry flavoured capsule to be red [1]. Colour can be used to increase the aesthetic value of products and reduce the threat of them being counterfeited. The development of unique colours for a particular active medicine, and the coloured printing can help to reduce the pitfalls of counterfeiting. [1] Tablets can be coloured by the addition of undoable colors to the tablet base(the basic ingredient of the capsule), or masking the natural colour of the active component and/ or other excipients. Sugar coating is considered to be the traditional system for coating tablets. Film coating involves the deposit of a thin polymer- grounded film onto the tablet core, generally by a spray system. [1] By their shape, size, and colour, capsules and tablets for a certain complaint can be categorized into tablets's common name such as Acetaminophen, Diphenhydramine, and Phenylephrine, as an illustration. Severe mislike and Phenylephrine hydrochloride of 5 mg/, diphenhydramine hydrochloride of 25 mg/, phenylephrine hydrochloride of 5 mg/, Sinus Headache acetaminophen 325 mg is the name of the blue capsule-shaped lozenge with the imprint 44 543. Woonsocket tradition Centre, Inc. provides it. [2] herbage, capsule-shaped lozenge with imprint 44 464 has been recognized as Acetaminophen 500 mg, Diphenhydramine12.5 mg, and Phenylephrine 5 mg for Allergy & Sinus Headache. It comes from Wal Mart StoresInc. [2] general name caffeine/ aspirin Bayer Headache Relief is a white, capsule- shaped lozenge with the mark Bayer Headache. 500 mg of aspirin and32.5 mg of caffeine. It comes from Bayer HealthCare LLC. [2]

III. CNN Methodology for Pharmaceutical Drug Classification

CNN (Convolutional Neural Network) is mainly useful and productive for the problems which needs image classification techniques. It achieves this by using a series of layers that are specifically designed to recognize different features in an image.

The CNN architecture contains many layers which has their own characteristics.

1) Input layer: This is where the raw pixel values of the image are input into the network.

2) Convolutional layer: With the help of some set of filters which can be learnable, this layer performs convolution operations on the image that we have given as input. Each filter scans the input image and produces a feature map that highlights a specific feature, such as edges, corners, or textures.

3) ReLU (Rectified Linear Unit) layer: ReLU layer name itself suggest that there is a use in ReLU activation function in this layer. The activation function is used to convolution layer to get the output. By applying the ReLU function the network can get the non-linearity characteristic and become more expensive.

4) Pooling layer: In this layer the process of down sampling of output of previous layer be occurs. This will be done by taking the maximum value in each sub region of the feature map. This helps to reduce the spatial dimensionality of the feature map and increase the efficiency of the network.

5) Fully connected layer: It is a layer which will take input as the flatten vector of the previous layer's output. And make the model to get trained for more complex representations of input.

6) Output layer: It gives the final output of the entire network. The produced output is one of the classes that image belong to.

The architecture of a CNN can be customized by increasing and decreasing the layers in contains, adjustment of the filter size and the number of neurons in each layer, and other hyperparameters. For the tasks, such as image classification, object detection, and semantic segmentation, there are different architectures been developed and discovered. [6,7,8,9]

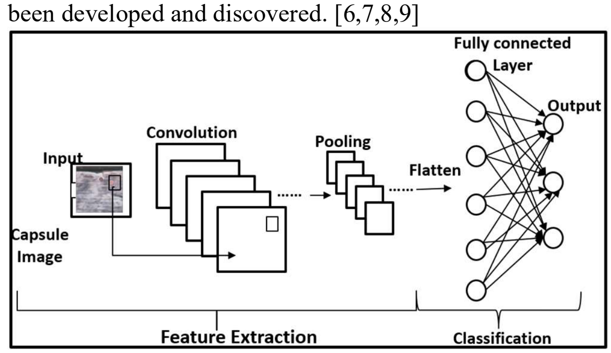

Fig.1. CNN Architecture for capsule identification

The CNN models used for this work are MobilenetV2.Densnet169, Densnet201, and Densnet121. Each model has a distinctive architecture.

Mobile Net V2:

The MobileNetV2 architecture consists of a series of layers that are optimized for efficient processing of image data. Here's a brief overview of the architecture:

1) Input layer: The input feeding layer for the model.
2) Convolutional layer: In this layer there is a 3x3 convolution operation which is applied to the input image to extract features.
3) Batch normalization layer: In this layer normalization of output of the convolutional layer will be done to reduce the effects of internal covariate shift.
4) ReLU activation layer: ReLU layer name itself suggest that there is a use in ReLU activation function in this layer. ReLU has been applied for getting non-linearity to the network.
5) Depth wise separable convolutional layer: In this layer there is a depth wise convolution be performed, which applies a single folter to each input channel. And after that a pointwise convolution which combines the outputs of the depth wise convolution with 1x1 filter. This decreases the number of parameters and computation while maintaining accuracy. Additional depth wise separable convolutional layers: The network repeats the depth wise separable convolutional layer multiple times to extract more complex features.
6) Linear bottleneck layer: This layer reduces the number of channels to improve efficiency.
7) Shortcut connection: In this layer the output of previous layer will be directly connected to the output of the output of the current layer. This allows gradients to flow directly to earlier layers.
8) Final layers: Finally at the end there is a a global average pooling layer and next to it another fully connected layer and a SoftMax activation layer, which produces the final output of the network.
Overall, MobileNetV2 is designed to balance accuracy with efficiency, making it well-suited for mobile and embedded vision applications.

DenseNet:

DenseNet-169 is another variant of the DenseNet architecture, which is designed to strike a balance between model size and performance. Here's an overview of the stages and transformations that an image of size 256x256 goes through in the DenseNet-169 architecture:
1) Initial convolutional layer: In this layer the image which given as input passed through a convolutional layer with a small kernel size to extract low-level features.
2) Dense blocks: The output that obtained from the initial convolutional layer is passed through a series of dense blocks, each of which consists of several densely connected convolutional layers. In DenseNet-169, there are 4 dense blocks, each with a different number of convolutional layers. The dense connections is useful for passing information |more easily through the network and reduce the risk of vanishing gradients.
3) Transition layers: Transition layers are inserted between the dense blocks to decrease the spatial dimensionality of the feature maps and helps to control the number of feature maps. There is a batch normalization layer, a 1x1 convolutional layer, and a pooling layer for these layer.
4) Final dense block: The final dense block is right after to global average pooling layer, which it helps averages the feature maps over the spatial dimensions. And after this there is a fully connected layer and a softmax activation layer, which produce the final output of the network.
Overall, DenseNet-169 is a slightly shallower and smaller variant of the DenseNet architecture that is designed to provide a good balance between accuracy and computational efficiency. It is suitable for applications where computational resources are limited but high accuracy is still required, such as mobile and embedded systems.

IV. EXPIRIMENTAL DESIGN AND PARAMETER SETTING

Internal mechanism for the paper, we have totally used 10 types of capsule images, which was trained and tested with 4 methods. Densnet architecture has one convolution layer, one pooling later, 3 dense block layers, 3 transition Layers and one classification layer.

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201
Convolution	112 X 112	7 X 7 conv,stride 2		
Pooling	56 X 56	3 X 3 max pool,stride 2		
Dense Block (1)	56 X 56	[1 X 1 conv, 3 X 3 conv] X 6	[1 X 1 conv, 3 X 3 conv] X 6	[1 X 1 conv, 3 X 3 conv] X 6
Transition Layer(1)	56 X 56	1 X 1 conv		
	28 X 28	2 X 2 average pool,stride 2		
Dense Block (2)	28 X 28	[1 X 1 conv, 3 X 3 conv] X 12	[1 X 1 conv, 3 X 3 conv] X 12	[1 X 1 conv, 3 X 3 conv] X 12
Transition Layer(2)	28 X 28	1 X 1 conv		
	14 X 14	2 X 2 average pool,stride 2		
Dense Block (3)	14 X 14	[1 X 1 conv, 3 X 3 conv] X 24	[1 X 1 conv, 3 X 3 conv] X 32	[1 X 1 conv, 3 X 3 conv] X 48
Transition Layer(3)	14 X 14	1 X 1 conv		
	7 X 7	2 X 2 average pool,stride 2		
Dense Block (4)	14 X 14	[1 X 1 conv, 3 X 3 conv] X 16	[1 X 1 conv, 3 X 3 conv] X 32	[1 X 1 conv, 3 X 3 conv] X 32
Classification Layer	1 X 1	7 X 7 global average pool		
		1000D fully-connected, softmax		

Fig.2**:** Densenet 121, 169, 201 architectures

DensNet121:

By the above table, we can understand the architecture which are used for ImageNet database. The stride which is for how many pixels are shifted across the input matrix. If stride parameter is set to n then it means that we have shifted n pixels at a time. DenseNet 121 consists of 121 layers, which are organized into four dense blocks and a final classification layer. Each dense block has a set of densely connected convolutional layers and a in the same block as input and gives a set of output feature maps. This densely connected architecture allows for feature reuse and facilitates gradient flow during training transition layer which reduces the spatial dimensionality of the feature maps.

Here's a brief overview of the internal layers and their role in DenseNet 121:

1) Convolutional Layers: The first layer in DenseNet 121 is a standard convolutional layer which takes image as input and applies a set of learnable filters to produce a set of feature maps.

2) Dense Blocks: The core of DenseNet 121 is the set of four dense blocks. Each of them contains various fully connected convolution layers that are arranged in a feedforward manner. Each layer in the dense block takes the feature maps produced by all previous layers in the same block as input and gives a set of output feature maps. This densely connected architecture allows for feature reuse and facilitates gradient flow during training.

3) Transition Layers: In this layer the process of reducing of spatial dimensionality of the feature maps produced by the dense blocks be take place. Each transition layer contains a batch normalization layer, a 1x1 convolutional layer, and a 2x2 average pooling layer. The 1x1 convolutional layer is used to reduce the number of feature maps and the average pooling layer decreases the spatial dimensionality of the feature maps.

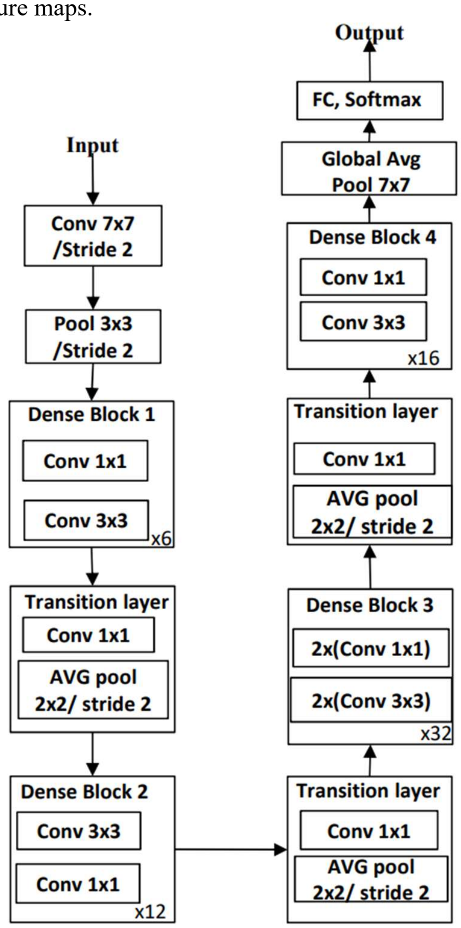

Fig.3. DenseNet 121 Architecture

4) Global Average Pooling: It is used to decrease the spatial dimensionality of the feature maps to a single value. This is done by taking the average of each feature map across its spatial dimensions.

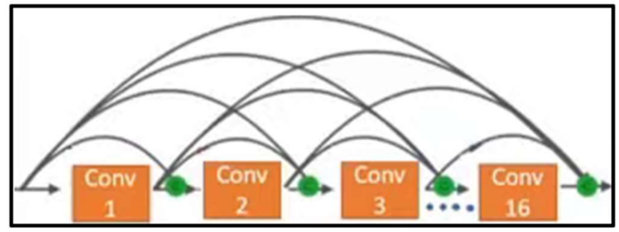

Fig.4. Dense Block in DenseNet 121 Architecture

5) Fully Connected Layer: It is the final layer in DenseNet 121, it is a fully connected layer which takes the output of the global average pooling layer as input and produces a set of class scores. The number of neurons in this layer represents to the number of classes in the classification problem.

Overall, DenseNet 121 has a highly parameter-efficient architecture that allows for strong feature reuse and effective gradient flow during training. This leads to high accuracy on image classification tasks with relatively small amounts of training data.

Densenet201:

DenseNet201 is a fully connected convolutional neural network architecture, which is an extension of the DenseNet121 architecture. It is designed to achieve even better performance on image classification tasks by increasing the depth of the network.

DenseNet201 has 201 layers, which are organized into four dense blocks and a final classification layer. Each dense block consists of several densely connected convolutional layers, and each layer in the dense block receives input from all previous layers in the same block.

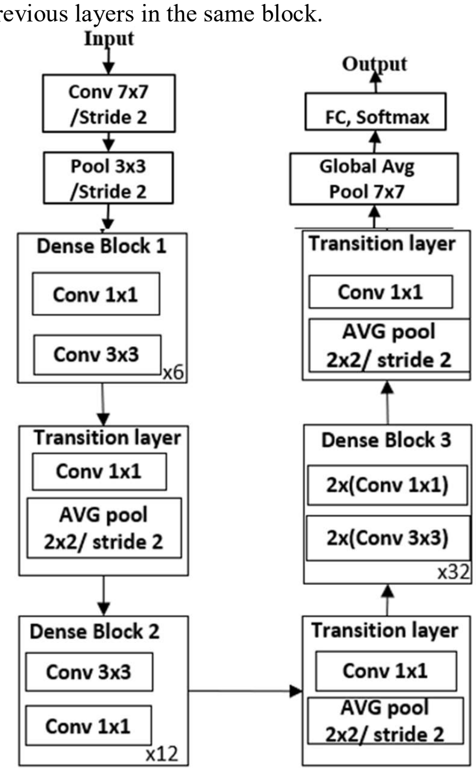

Fig.5: DenseNet 201 Architecture

Here's an overview of the internal layers of DenseNet201:

1) Convolutional Layers: The first layer in DenseNet201 is a standard convolutional layer which takes the image as input

and applies a set of learnable filters to produce a set of feature maps.

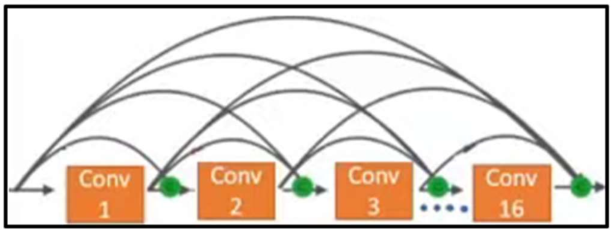

Fig.6. Dense Block in DenseNet 201 Architecture

2) Dense Blocks: The core of DenseNet201 is the set of four dense blocks, which are similar to those in DenseNet121. Each dense block consists of multiple fully connected convolutional layers. The output of every layer is linked with the input of all subsequent layers in the same block. This dense connectivity allows the network to reuse features and encourages feature propagation.
3) Transition Layers: In this layer spatial dimensionality of the feature maps produced by the dense blocks will be reduced. Each transition layer will contains a batch normalization layer, a 1x1 convolutional layer, and a 2x2 average pooling layer. The 1x1 convolutional layer is used to reduce the number of feature maps, while the average pooling layer reduces the spatial dimensionality of the feature maps.
4) Global Average Pooling: This is the layer which helps to decrease the spatial dimensionality of the feature maps to a single value. This is done by taking the average of each feature map across its spatial dimensions.
5) Fully Connected Layer: It is the final layer in DenseNet201 which is a fully connected layer. It takes the output of the global average pooling layer as input and produces a set of class scores. The number of neurons in this layer is same as the number of classes in the classification problem.

MobileNetV2:
MobileNetV2 is a popular CNN architecture for mobile devices that was proposed by Sandler et al. in 2018. It is designed to achieve high accuracy on various computer vision tasks while being optimized for low latency and low power consumption. The MobileNetV2 architecture consists of several building blocks, including depth wise separable convolutions and linear bottlenecks, which help to reduce the number of parameters and computational complexity of the network. Here's an overview of the internal layers of MobileNetV2.
1) Convolutional Layers: The first layer in MobileNetV2 is a standard convolutional layer that takes the input image and applies a set of learnable filters to produce a set of feature maps.
2) Inverted Residual Blocks: The core building blocks of MobileNetV2 are the inverted residual blocks. Each block consists of three layers: one 1x1 convolutional layer, one depth wise separable convolutional layer, and another 1x1 convolutional layer. The first 1x1 convolutional layer is used to increase the number of channels, while the depth wise separable convolutional layer performs the actual convolutions. The second 1x1 convolutional layer is used to decrease the number of channels which is back to the original number.
3) Linear Bottlenecks: The inverted residual blocks in MobileNetV2 use linear bottlenecks, which are 1x1 convolutions with a linear activation function. This helps to decrease the number of parameters and computational complexity of the network while preserving accuracy.
4) Expansion Layers: Inverted residual blocks may also include expansion layers, which are 1x1 convolutional layers that are helps to increase the number of channels before the depth wise separable convolutional layer
5) Batch Normalization: MobileNetV2 also uses batch normalization after each convolutional layer for improving the convergence of the network and to reduce overfitting.
6) Global Average Pooling: This is the layer which is used to decrease the spatial dimensionality of the feature maps to a single value. This is done by taking the average of every feature map across its spatial dimensions.
7) Fully Connected Layer: This is the final layer in MobileNetV2 and it is a fully connected layer. It takes the output of the global average pooling layer as input and produces a set of class scores. The number of neurons in this layer is as similar as the number of classes in the classification problem

V. Results and Discussion

DenseNet121:
In this work, densenet121 provided an accuracy of 81.52%. Since the test loss value is so near to 0, which is why it give the best results for the model with the test loss value was 0.74594.

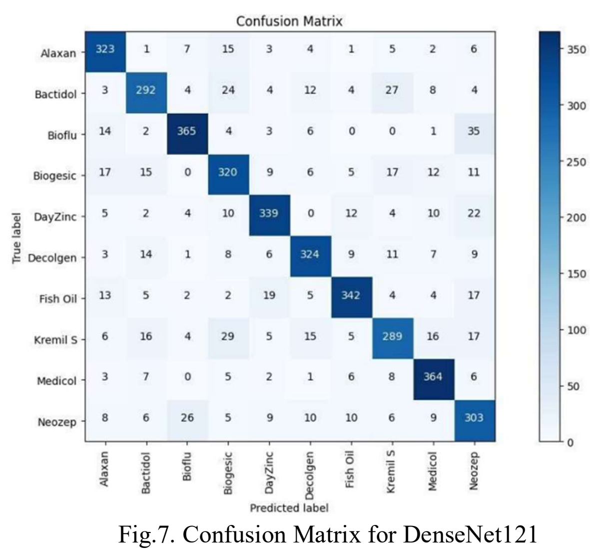

Fig.7. Confusion Matrix for DenseNet121

The confusion matrix shows that the ‘Medicol’ and ‘Bioflu’ were classified most accurately and ‘Kremil S’ had the most errors.

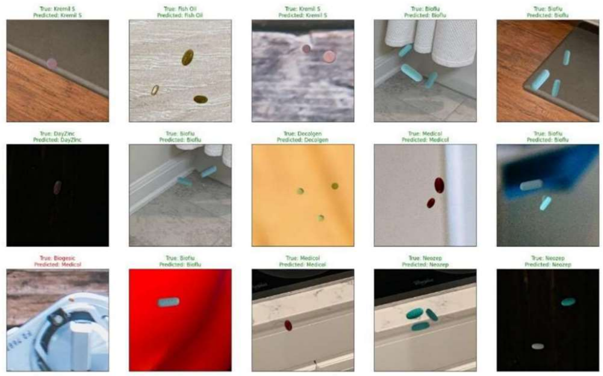

Fig.8. DenseNet121 predisction-1

A sample of 15 images were randomly tested and the densenet121 has given 14 correct classifications.

	Precision	Recall	F1-score	Support
Alaxan	0.82	0.88	0.85	367
Bactidol	0.81	0.76	0.79	382
Bioflu	0.88	0.85	0.87	430
Biogesic	0.76	0.78	0.77	412
DayZinc	0.85	0.83	0.84	408
Decolgen	0.85	0.83	0.84	392
Fish Oil	0.87	0.83	0.85	413
Kremil S	0.78	0.72	0.75	402
Medicol	0.84	0.91	0.87	402
Neozep	0.70	0.77	0.74	392
Accuracy	-	-	0.82	4000
Macro Avg	0.82	0.82	0.81	4000
Weighted Avg	0.82	0.82	0.82	4000

Fig.9. DenseNet121 predisction-2

In Fig-9, the obtained values of respective classes have been shown in the report and its very clear that the accuracy precision and F1 score and support have almost similar values.

Graphs:

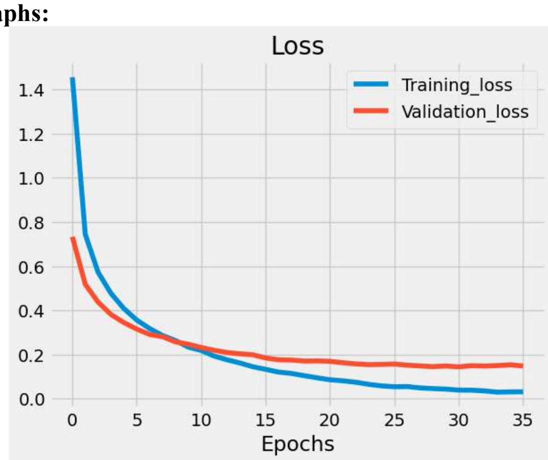

Fig.10. DenseNet121 training loss graph

This is the test loss graph of Dense Net 121 architecture. The training loss of the model is high at the beginning of the epochs and gradually decreased and reached near to zero at 35th epoch. The training model has highest loss at 1 and 2 epochs. And in the graph the validation loss of the model has high values which means it has high validation loss at the beginning epochs.

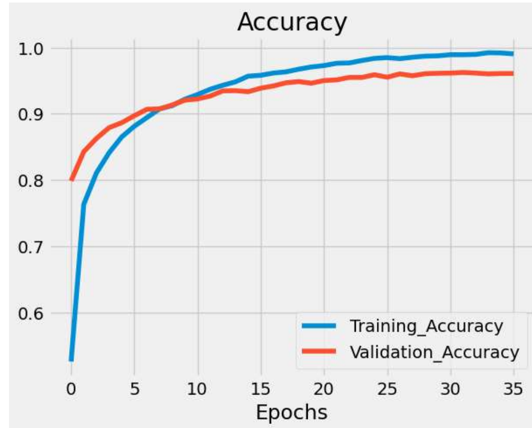

Fig.11. DenseNet121 Training Accuracy graph

In fig-11, it clearly mentioning that until 20 epochs the training loss decreased very rapidly and then there is a little change in the validation loss and training loss was continuously declining until 160 epochs, Similarly, the training and validation accuracy increased sharply until 20 epochs and then there is a little increase in every epoch after that.

DenseNet201:

For this application, densenet121 provided an accuracy of 96.10%. Since the test loss value is so near to 0, we were able to get the best results for the model with the test loss value was 0.12955.

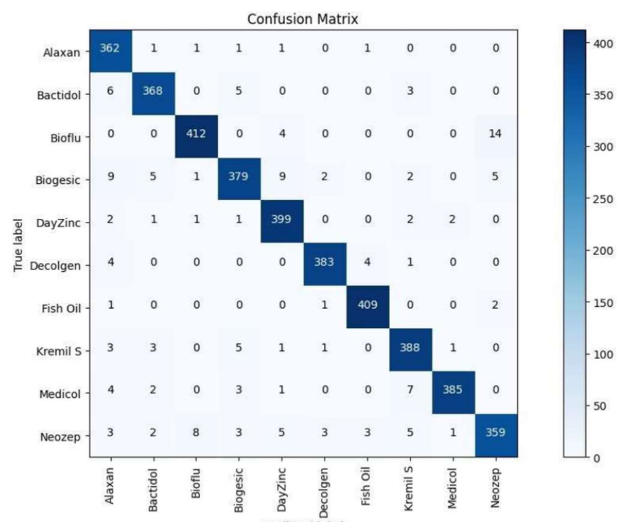

Fig.12. Confusion Matrix for DenseNet201

The confusion matrix shows that the 'Bioflu' and 'Fish Oil' were classified most accurately and 'Neozep' had the most errors.

	Precision	Recall	F1-score	Support
Alaxan	0.92	0.99	0.95	367
Bactidol	0.96	0.96	0.96	382
Bioflu	0.97	0.96	0.97	430
Biogesic	0.95	0.92	0.94	412
DayZinc	0.95	0.98	0.96	408
Decolgen	0.98	0.98	0.98	392
Fish Oil	0.98	0.99	0.99	413
Kremil S	0.95	0.97	0.96	402
Medicol	0.99	0.96	0.97	402
Neozep	0.94	0.92	0.93	392
Accuracy	-	-	0.96	4000
Macro Avg	0.96	0.96	0.96	4000
Weighted Avg	0.96	0.96	0.96	4000

Fig.13. DenseNet201 predisction-2

The above figure shows that the records of the prediction of various capsules that have been trained. Its already known that recall is the ability of a model to classify and identify all the data points in a relevant class. Here for densenet201 recall of the fish oil was high which means that tablet was identified accurately throughout all other relevant classes. Precision of a model is the ability of a model to classify and identify the only data points in same class.

Graphs:

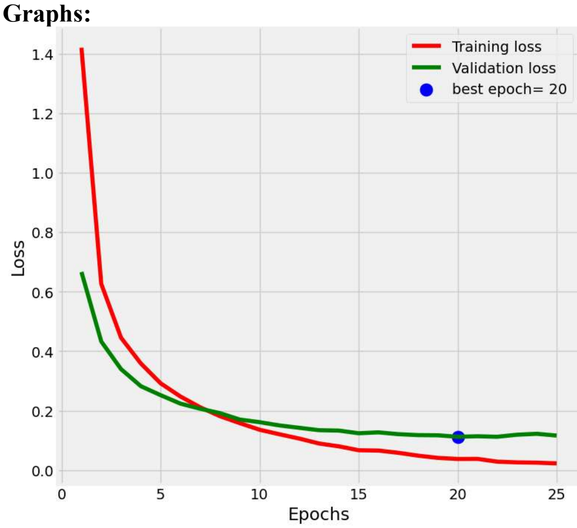

Fig.14. DenseNet201 training loss graph

The above graphs show that the training accuracy of the model is 1.4 at the beginning and got decreased to nearly 0.1 at 20th epoch. And the validation loss of the model is stared at 0.7 and got decreased to o.1 at 20th epoch which results 20 is the best epoch for the model.

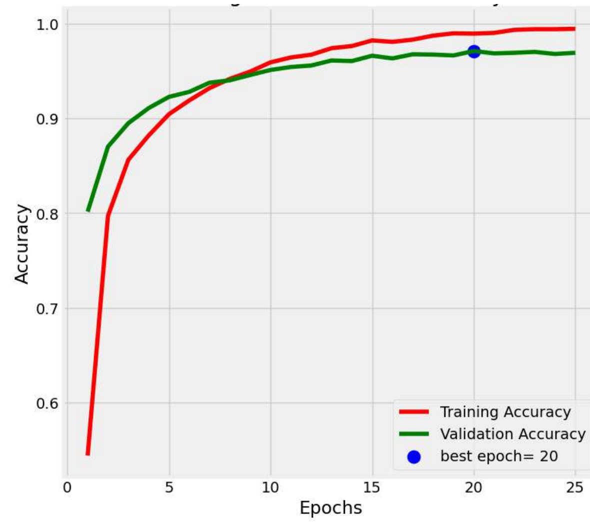

Fig.15. DenseNet201 training accuracy graph

The above graphs show that the training accuracy of the model is 3% at the beginning epochs and got increased to nearly 99% at 20th epoch. And the validation accuracy of the model is stared with 80% and got increased to nearly 96% at 20th epoch, which results 20 is the best epoch for the model.

DenseNet169:

For this work, densenet169 provided an accuracy of 96.22%. Since the test loss value is so near to 0, that's why the model got best results, where the test loss value was 0.13251.

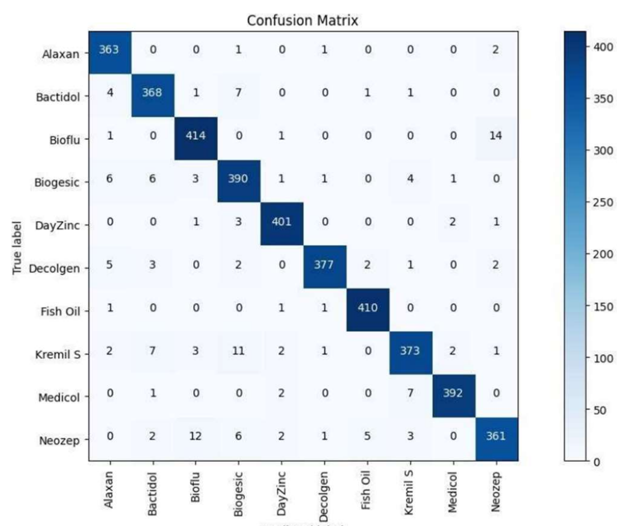

Fig.16. Confusion Matrix for DenseNet169

For DenseNet169, by the above confusion matrix say that 'Bioflu' and 'Fish Oil' were classified most accurately and 'Neozep' had the most errors.

	Precision	Recall	F1-score	Support
Alaxan	0.95	0.99	0.97	367
Bactidol	0.95	0.96	0.96	382
Bioflu	0.95	0.96	0.96	430
Biogesic	0.95	0.95	0.94	412
DayZinc	0.95	0.98	0.98	408
Decolgen	0.95	0.96	0.97	392
Fish Oil	0.95	0.99	0.99	413
Kremil S	0.95	0.93	0.94	402
Medicol	0.95	0.98	0.98	402
Neozep	0.95	0.92	0.93	392
Accuracy	-	-	0.96	4000
Macro Avg	0.95	0.96	0.96	4000
Weighted Avg	0.95	0.96	0.96	4000

Fig.17. DenseNet169 predisction-2

By the prediction of DenseNet169, it shows that the Precision of every tablet is same and there are variations in the recall and f1-score. 'Alaxin' and 'Fish Oil' got high F1-score which means they have identified most accurately in same class and in some relevant classes too.

Graphs:

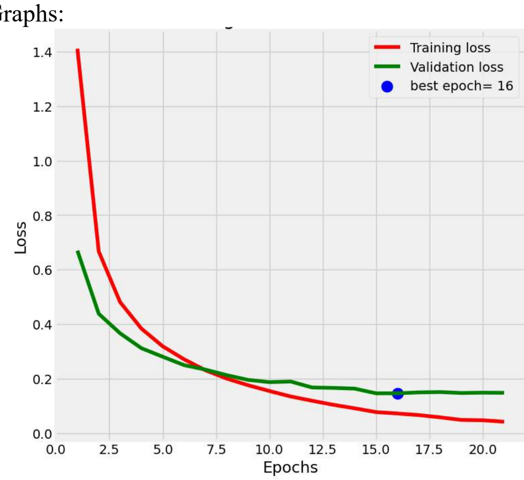

Fig.18. DenseNet169 training loss graph

At first epoch the value of training loss is 1.4 and gradually it started decreasing whereas the val_loss value at point 0 is around 0.7 and eventually it starts decreasing. The loss got reduced nearly to zero at 30^{th} epoch. And the above graph shows that at 16^{th} epoch, the loss of the model is very less which says it is the best epoch.

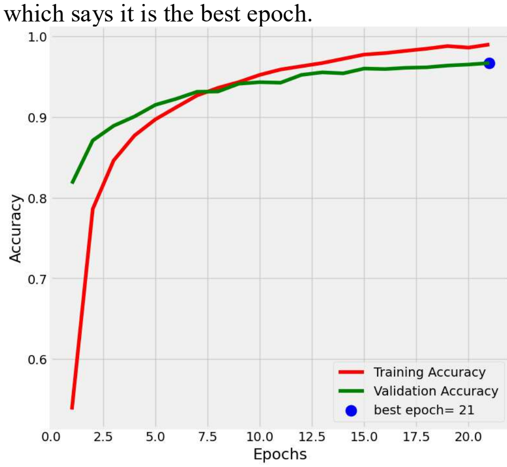

Fig.19. DenseNet169 training accuracy graph

For DenseNet169, the training accuracy gradually increased and reached mostly 99 to 99 percent. Similarly, the validation accuracy of the model at the starting epochs it was there around 83 percent and got reached 94 to 96 percent when it completes 30 epochs. And as we can see that at 21^{h} epoch, we got highest accuracy. So, we can say that it is the best epoch.

MobileNetV2:

In this work, densenet169 provided an accuracy of 84.35%. Since the test loss value is so near to 0, we were able to get the best results for the model with the test loss value was 0.47399.

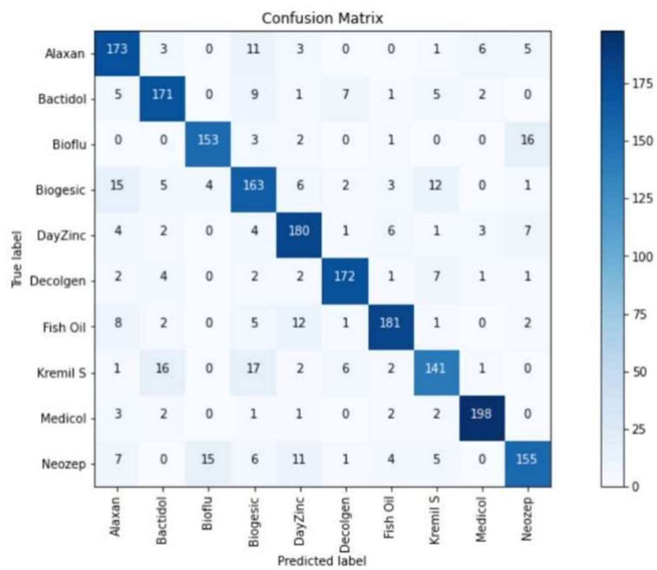

Fig.20. Confusion Matrix for MobileNetV2

For MobileNetV2, the above confusion matrix says that 'Medicol' and 'Fish Oil' were classified accurately compared to other capsules and 'Kremil S' had errors.

	Precision	Recall	F1-score	Support
Alaxan	0.79	0.86	0.84	202
Bactidol	0.83	0.85	0.84	201
Bioflu	0.89	0.87	0.88	175
Biogesic	0.74	0.77	0.75	211
DayZinc	0.82	0.87	0.84	208
Decolgen	0.91	0.90	0.90	192
Fish Oil	0.90	0.85	0.88	212
Kremil S	0.81	0.76	0.78	186
Medicol	0.94	0.95	0.94	209
Neozep	0.83	0.76	0.79	204
Accuracy	-	-	0.84	2000
Macro Avg	0.85	0.84	0.84	2000
Weighted Avg	0.84	0.84	0.84	2000

Fig.21. MobileNetV2 Classification report

By the prediction of MobileNetV2, it shows that this model predicted somehow less accurately when compared to the Dense Net architecture. The table clearly defines that the precision and recall scores of the capsules, 'medicol' got nearly 94 percent and 95 percent respectively.

Graphs:

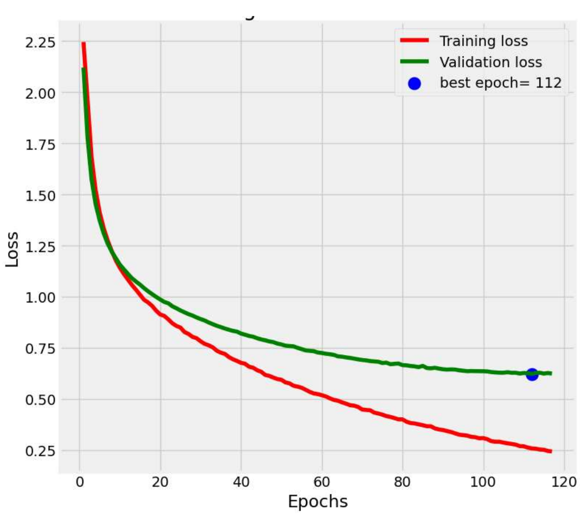

Fig.22. MobileNetV2 training loss graph

The loss in MobileNetV2 is nearly similar to all other models, high at the starting epochs and gradually decreased nearly to 0.25 at 112th epoch. At the same time validation loss of the model started at 2.25, got decreased and reached 0.6 in 112th epoch. This says that say that 112 is the best epoch of the model.

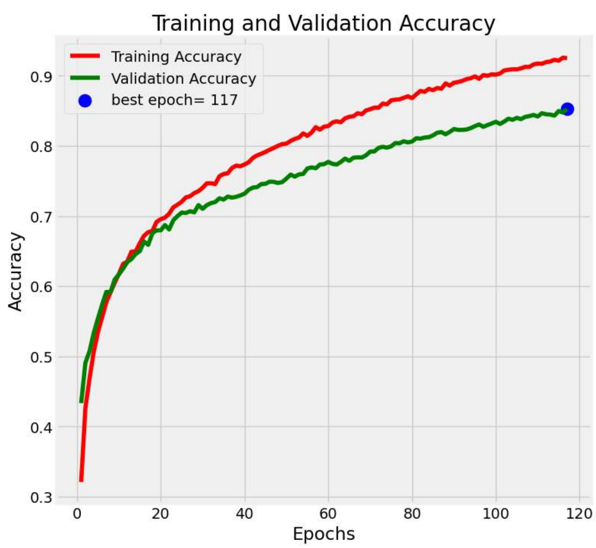

Fig.23. MobileNetV2 training accuracy graph

The Training accuracy of this model reached nearly 95 percent at the 117th epoch and validation accuracy got reached nearly 85 percent. This says that 117 is the best epoch for this model.

Performance Comparison:

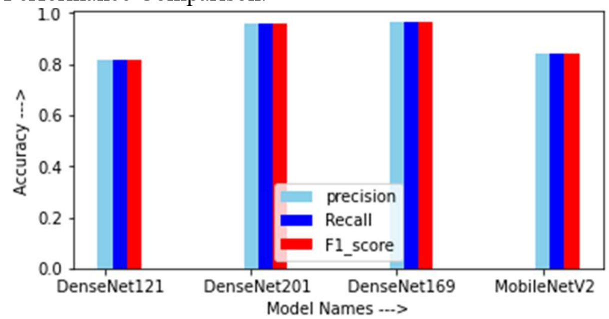

Fig.24. Performance Comparison

Here are the performance comparision of this work, there are 4 models in total they are DenseNet121, DenseNet201, DenseNet169 and MobleNetV2. The overall accuracy of the models DenseNet121, DenseNet201, DenseNet169 and MobleNetV2 are 81.5%, 96.1%, 96.2 and 84.3 respectively.

VI. Conclusion

By this work reader could get an idea of how deep learning is useful for image classification. The methods that this work used here are some of the best techniques for image classification. There is a suggestion for the readers, to use these methods to train data and get the accurate results. Out of these four architectures DenseNet169 and DenseNet201 has got highest accuracy of 96.2% and 96.1% respectively, which reflect that these are the best method to use.

References

[1] Rozenwald, M. B., Galitsyna, A. A., Sapunov, G. V., Khrameeva, E. E., & Gelfand, M. S. (2020). A machine learning framework for the prediction of chromatin folding in Drosophila using epigenetic features. *PeerJ Computer Science*, *6*, e307.

[2] Pundir, S., & Verma, A. M. L. (2014). Review on lozenges. *Journal der pharmazie Forschung*, *2*(1), 1-10.

[3] Chasanah, U., Yulian, F., Yulianty, R., Irianto, A., & Radjaran, A. (2017, October). The Effects of Binder on the Physical Characteristic of Squeezed Rhizome of Kaempferia galanga L Lozenges. In *Health Science International Conference (HSIC 2017)* (pp. 264-270). Atlantis Press.

[4] Firriolo, J. F. (1994). Oral cavity-A Review. *Oral Surg Med Oral Pathol*, *78*(2), 189-93.

[5] Batheja, P., Thakur, R., & Michniak-Kohn, B. (2006). Basic biopharmaceutics of buccal and sublingual absorption. In *Enhancement in drug delivery* (pp. 176-202). CRC Press.

[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

[7] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*, 211-252.

[8] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

[9] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

[10] Huang, G., Liu, Z., & Maaten, L. (2016). van, and Weinberger. *Kilian Q,"Densely Connected Convolutional Networks," arXiv. org*.