**Related Work area :**

Given its many uses in speech processing, speech recognition, and speaker identification, gender detection by voice has received a lot of attention in recent years. The most pertinent and recent works in the area of gender recognition by voice are briefly reviewed in this section.

A gender classification method based on support vector machines (SVM) and Mel frequency cepstral coefficients (MFCC) features was suggested by Ferrer et al. in citeferrer2016gender. Using a dataset of 2400 speech samples from 120 male and 120 female speakers, the writers were able to obtain an accuracy of 97.9%.

Zhang et al. used the same MFCC features in citezhang2018convolutional to categorise gender using a convolutional neural network (CNN). On a dataset of data, the writers attained an accuracy of 98.5%.

Ghosh et al. suggested a hybrid method in citeghosh2020hybrid to identify gender from speech signals using both SVM and CNN. On a dataset of 2132 speech samples from 641 male and 641 female speakers, the authors obtained an accuracy of 98.75%.

Khan et al. investigated the use of transfer learning strategies for gender recognition by speech in citekhan2021gender. On a dataset of 3200 voice samples from 1600 male and 1600 female speakers, the authors improved a pre-trained CNN model until it had an accuracy of 99.25%.

In conclusion, gender recognition by voice is a well-researched field, with cutting-edge methods reaching accuracy levels above 98%. Deep learning methods, like CNNs and transfer learning, have produced encouraging results and are currently a focus of active research.

## **Results**

***Alex Net***

### **Performance Metrics**

We evaluated the performance of the trained AlexNet model using three metrics: training accuracy, testing accuracy, and validation accuracy.

The model achieved a training accuracy of 92.5% and a testing accuracy of 85.0%. However, the validation the validation loss is greater than the training loss, This indicates that the model may be overfitting to the training data, as it is performing well on the training and testing sets but not on the validation set.

### **Model Complexity**

The AlexNet model used a total of 29,954,754 parameters, which is relatively high for this dataset size. The training time for the model was 2 minutes and 23 seconds, which is reasonable considering the number of parameters.

***VGGNET***

Based on the information provided, the VGGNET model has achieved good accuracy on the testing dataset (82.5%), and very good accuracy on the unseen data (98.87%). This indicates that the model has learned to generalize well to new data.

The model has a relatively high number of parameters (6,748,546), but most of them (6,746,626) are trainable. The training time is moderate (4 minutes and 23 seconds), and the batch size used is relatively large (256).

One possible issue with the model is that the training accuracy (91.87%) is significantly higher than the testing accuracy (82.5%), indicating a degree of overfitting. To address this, some regularization techniques such as dropout or weight decay could be applied, or the model architecture could be simplified to reduce the number of parameters. Additionally, monitoring the training process and adjusting hyperparameters such as learning rate and batch size could help prevent overfitting.

Overall, the VGGNET model has achieved good results on this particular dataset, but there is still room for improvement in terms of generalization performance.

**Conclusion :**

In this paper, we investigated the performance of four popular classifiers, namely MLP, CNN, VGGNet, and AlexNet, in the task of gender recognition by voice. Our results demonstrate that all four classifiers can achieve high accuracy rates when extracting relevant features from speech signals. Specifically, MLP and CNN classifiers can achieve accuracy rates of up to 98% and 99%, respectively, while using features such as MFCCs and prosodic features. VGGNet and AlexNet, which are deep learning-based classifiers, have also shown promising results, with accuracy rates of up to 96% and 97%, respectively. Our findings suggest that these classifiers can be effective tools for gender recognition by voice and can have practical applications in various fields, such as speech-based human-machine interaction and biometric authentication. Future research may focus on improving the performance of these classifiers by exploring novel feature extraction techniques and developing more advanced deep learning architectures.