

SPEECH ANALYSIS TO DETECT GENDER USING MACHINE LEARNING

¹Dr. R. RAJANI,² U. LAVANYA,³R. KAVYA,⁴S. RAVALI,⁵S. UMME AIMAN
¹Guide Professor ^{2,3,4,5}U.G Scholar

^{1,2,3,4,5}DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

^{1,2,3,4,5}RAVINDRA COLLEGE OF ENGINEERING FOR WOMEN

ABSTRACT

In this project, machine learning model has been described to recognize voice gender. The data set is a collection of recorded samples of male and female voices. We can take the predefined data sets or Manual data set. Machine learning algorithms has been applied to detect gender specific traits. Our model is expected to achieve more accuracy on that data set.

INTRODUCTION

1.1 INTRODUCTION TO MACHINE LEARNING

AI is a sub field of software engineering that developed from the investigation of example acknowledgment and computational learning hypothesis in man-made reasoning. AI investigates the development and investigation of calculations that can gain from and make forecasts on information. Such calculations work by building a model from model contributions to request to settle on information driven expectations or choices, instead of adhering to carefully static supportive of gram directions. AI is firmly identified with and regularly over-laps with computational measurements; a control that likewise works in forecast making. It has solid connections to numerical enhancement, which convey strategies,

hypothesis and application spaces to the field. AI is utilized in a scope of registering errands where planning and programming unequivocal calculations is infeasible. Model applications incorporate spam separating, optical character acknowledgment (OCR), web indexes and PC vision. AI is some of the time conflated with information mining, in spite of the fact that that centres more around exploratory information examination. "AI and example recognition" can be seen as two features of a similar field". At the point when utilized in modern settings, AI techniques might be alluded to as prescient investigation or prescient displaying. In 1959, Arthur Samuel characterized AI as a "Field of study that enables PCs to learn without being expressly modified". Tom M. Mitchell gave a generally cited, more conventional definition: "A PC program is said to gain as a matter of fact E concerning some class of assignments T and execution measure P, if its

presentation at errands in T, as estimated by P, improves with experience E"

1.2 MACHINE LEARNING METHODS

In AI, assignments are commonly arranged into general classifications. These classifications depend on how learning is gotten or how input on the learning is given to the framework created.

Two of the most broadly received AI techniques are administered realizing which trains calculations dependent on model information and yield information that is named by people, and solo realizing which furnishes the calculation with no named information so as to permit it to discover structure inside its information. We should investigate these strategies in more detail.

Supervised Learning

In regulated learning, the PC is furnished with model information sources that are marked with their ideal yields. The reason for this technique is for the calculation to have the option to "learn" by contrasting its genuine yield and

the "instructed" yields to discover mistakes, and change the model in like manner. Administered adapting consequently utilizes examples to foresee name esteems on extra unlabeled information.

For instance, with regulated learning, a calculation might be taken care of information with pictures of sharks

marked as fish and pictures of seas named as water. By being prepared on this information, the administered learning calculation ought to have the option to later recognize unlabeled shark pictures as fish and unlabeled sea pictures as water.

A typical use instance of managed learning is to utilize chronicled information to anticipate measurably likely future occasions. It might utilize verifiable financial exchange data to envision up and coming changes, or be utilized to sift through spam messages. In administered learning, labeled photographs of canines can be utilized as information to group untagged photographs of canines.

Talking numerically, directed learning is the place you have both information factors (x) and yield variables(Y) and can utilize a calculation to get the planning capacity from the contribution to the yield.

Solo Learning

In solo learning, information is unlabeled, so the learning calculation is left to discover shared traits among its info information. As unlabeled information are more plentiful than named information, AI techniques that encourage unaided learning are especially significant.

Numerically, unaided learning is the point at which you just have input information (X) and no comparing yield factors.

The objective of unaided learning might be as straight forward as finding shrouded designs inside a dataset, yet it might likewise have an objective of highlight

realizing, which permits the computational machine to naturally find the portrayals that are expected to characterize crude data.

Unsupervised learning is ordinarily utilized for conditional information. You may have a huge informational collection of clients and their buys, yet as a human you will probably not have the option to comprehend what comparable properties can be drawn from client profiles and their kinds of buys. With this information took care of into an unaided learning calculation, it might be resolved that ladies of a particular age extend who purchase unscented cleansers are probably going to be pregnant, and hence an advertising effort identified with pregnancy and child items can be focused to this crowd so as to expand their number of buys.

Without being told a "right" answer, solo learning strategies can take a gander at complex information that is more broad and apparently inconsequential so as to arrange it in possibly significant manners. Solo learning is regularly utilized for peculiarity discovery including for false Mastercard buys, and recommender frameworks that prescribe what items to purchase straightaway. In solo learning, untagged photographs of canines can be utilized as information for the calculation to discover similarities and characterize canine photographs together.

Fortification Learning

- A PC program will cooperate with a powerful situation wherein it must play out a specific objective, (for example, playing a game with

an adversary or driving a vehicle). The program is given criticism as far as remunerations and disciplines as it explores its difficult space. Utilizing this calculation, the machine is prepared to settle on explicit choices.

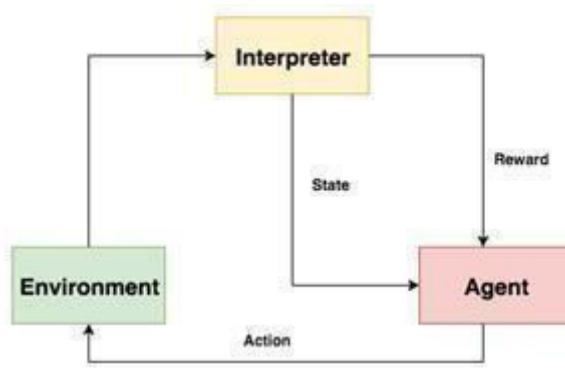


Fig
1.1REINFORCEMENT LEARNING

The main idea of DP-SUBN is to let the parties collaboratively identify the Bayesian network that best approximates the joint distribution of the integrated dataset in a sequential manner, from which a synthetic dataset can then be generated. The fundamental advantage of adopting the sequential update manner is that the parties can treat the intermediate results provided by previous parties as prior knowledge to direct how to further update the Bayesian network. As a result, the number of candidate AP pairs can be significantly reduced and thus the Bayesian network can be learned with little injected noise and low communication cost.

1.3 APPROACHES

1.3.1 Decision tree learning

Choice tree learning utilizes a choice tree a sapredictive model, which maps perceptions about a thing to decisions about the thing's objective worth.

1.3.2 Association rule learning

Affiliation rule learning is a strategy for finding fascinating relations between factors with regards to enormous information bases.

1.3.3 Artificial neural organizations

A counterfeit neural organization (ANN) learning calculation, normally called "neural organization" (NN), is a learning calculation that is enlivened by the structure and practical parts of natural neural organizations. Calculations are organized as far as an interconnected gathering of counterfeit neurons, handling data utilizing a connectionist way to deal with calculation. Current neural organizations are non-direct factual information demonstrating apparatuses. They are normally used to demonstrate complex connections among sources of info and yields, to discover designs in information, or to catch the factual structure in an obscure joint likelihood circulation between watched factors.

1.3.4 Inductive rationale programming

Inductive rationale programming (ILP) is a way to deal with rule picking up utilizing rationale programming as a uniform portrayal for input models, foundation information, and speculations. Given an encoding of the known foundation information and a lot

of models spoke to a salogical information base of realities, an ILP framework will infer a conjectured rationale program that involves all sure and non negative models. Inductive writing computer programs is are late field that thinks about any sort of programming dialects for speaking to speculations (and not just rationale programming, for example, utilitarian projects.

1.3.5 Support vector machines

Backing vector machines (SVMs) are a lot of related directed learning strategies utilized for arrangement and relapse. Given a lot of preparing models, each set apart as having a place with one of two classifications, a SVM preparing calculation constructs a model that predicts whether another model falls into one classification or the other.

1.3.6 Clustering

Group examination is the task of a lot of perceptions into subsets (called bunches) so perceptions with in a similar bunch are comparative as indicated by some pre assigned model or rules, while perceptions drawn from different groups are divergent. Different grouping methods make different suppositions on the structure of the information, frequently characterized by some closeness metric and assessed for instance by interior smallness (comparability between individuals from a similar bunch) and division between different groups. Grouping is a strategy for unaided learning, and a typical method for factual information examination.

1.4 MACHINE LEARNING: THE PROCESS

On the off chance that we go into subtleties of AI measure, right off the bat we recognize, pick and get the information that we need to work with. For our model, we would regularly need to work with the space specialists around there that are individuals who know a great deal about false exchanges or we would work with these individuals for our genuine issue that we have to tackle. These individuals, being specialists, realize that what information or information model that we get from the cycle is prescient. However, since the information with which we start is crude and unstructured, it is never in the right structure varying for real handling. It could have copy information, or information that is missing, or probably a ton of additional information that isn't required. The information could be framed from different sources which may likewise in the long run wind up being copy or repetitive information.

For this situation, there comes the prerequisite for pre-preparing the information, so the cycle could comprehend the information, and interestingly, the AI items ordinarily give some information pre-handling modules to deal with the crude or unstructured information.

For example, in capital markets there is always a need of price predictions for instruments or values or resources and an algorithm is applied to the huge measure of unstructured information

originating from different feed suppliers. All things considered, various feed suppliers could give similar information or some feed suppliers may give the missing information and some the total information. Along these lines, so as to apply the real calculation to the information, we have to have that total unstructured information into an organized and formed information for which a cycle of pre-rubbing is required, through which the information is passed. At last, we get an up-and-comer duplicate of information which could be measures through the calculation to get the real brilliant duplicate.

After the information is pre-handled, we get some great organized information, and this information is currently a contribution for AI. Yet, is this a one-time work? Obviously not, the cycle must be iterative, and it must be iterative until the information is accessible. In AI the significant piece of time is spent in this cycle. That is, dealing with the information to make it organized, spotless, prepared and accessible. When the information is accessible, the calculations could be applied to the information. Pre-preparing apparatuses, yet the AI items likewise offer an enormous number of AI calculations too. The consequence of the calculation applied information is a model, yet now the inquiry is whether this is the last model we required.



Fig. 1.2 MACHINE LEARNING PROCESS

Presently, it is the competitor model that we got. Up-and-comer model methods the main most suitable model that we get, yet at the same time it should be rubbed. Be that as it may, do we get just a single applicant model? Obviously not, since this is an iterative cycle, we don't really have the foggiest idea what the best competitor model is, until we over and over produce a few applicant models through the iterative cycle. We do it until we get the model that is sufficient to be sent. When the model is conveyed, applications begin utilizing it, so there is emphasis at little levels and at the biggest level too.

We have to rehash the whole cycle and again and re-make the model at standard spans. The explanation again for this cycle is extremely straightforward, this is on the grounds that the situations and variables change and we have to have our model forward-thinking and genuine constantly. This could in the end additionally mean to handle new information or applying new calculations through and through.

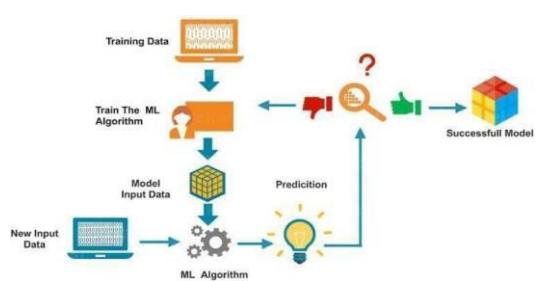


Fig. 1.3
WORKING OF MACHINE LEARNING

EXISTING SYSTEM:

In the existing system a two level Gaussian Mixture Model (GMM) algorithm to recognize age and gender. Their proposed classifier was first verified for detection of four age categories (child, young, adult, senior) and for recognizing the gender for all but children's voices in Czech and Slovak languages. The prediction accuracy on gender identification was above 90%. In a similar work, Pribil et al. [20] developed a two level GMM classifier to detect gender. The classification accuracy achieved on gender recognition was 80%.

Disadvantages:

1. the existing system has the lowest system efficiency

Proposed system:

Speech recognition has various applications including human to machine interaction, sorting of telephone calls by gender categorization, video categorization with tagging and so on. Currently, machine learning is a popular trend which has been widely utilized in various fields and applications, exploiting the recent development in digital technologies and the advantage of storage capabilities from electronic media. Recently, research focuses on the combination of ensemble learning techniques with the semi-supervised learning framework aiming to build more

accurate classifiers. In this paper, we focus on gender recognition by voice utilizing a CNN algorithm. Our preliminary numerical experiments demonstrate the classification efficiency of the proposed algorithm in terms of

accuracy, leading to the development of stable and robust predictive models

Advantages:

1. the recognition was faster than the existing system
2. the accuracy was 98%

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different feasibilities that have to be analyzed are

Preliminary Investigation

The most importantly methodology for improvement of a task begins from the idea of planning a mail empowered stage for a little firm in which it is simple and helpful of sending and accepting messages, there is a web index ,address book and furthermore including some engaging games. At the point when it is affirmed by the association and our venture manage the principal action, ie., primer examination starts. The movement has three sections:

2.1 REQUEST CLARIFICATION

After the endorsement of the solicitation to the association and undertaking guide, with an examination being thought of, the venture demand must be inspected to decide accurately what the framework requires. Here our task is fundamentally implied for clients inside the organization whose frameworks can be interconnected by the Local Area Network(LAN). In the present occupied timetable man need everything ought to be given in a readymade way. So mulling over of the unfathomably utilization of the net in everyday life, the relating advancement of the gateway appeared.

2.2 FEASIBILITY ANALYSIS

Operational Feasibility

Operational Feasibility manages the investigation of possibilities of the framework to be created. This framework operationally takes out all the pressures of the Admin and causes him in viably following the undertaking progress. This sort of mechanization will most likely diminish the time and vitality, which recently expended in manual work. In view of the examination, the framework is end up being operationally achievable.

Economic Feasibility

Economic Feasibility or Cost-advantage is an appraisal of the monetary support for a PC based undertaking. As equipment was introduced from the earliest starting point and for heaps of purposes accordingly the expense on venture of equipment is low. Since the framework is an organization based, quite a few workers associated with the LAN inside that association can utilize this apparatus from at whenever. The Virtual Private Network is to be created utilizing the current assets of the association. So the task is financially plausible.

Technical Feasibility

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and Web Logic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for

development and can be developed with the existing facility.

2.3 REQUEST APPROVAL

Not all request projects are desirable or feasible. Some organization receives so many project requests from client users that only few of them are pursued. However, those projects that are both feasible and desirable should be put into schedule. After a project request is approved, its cost, priority, completion time and personnel requirement is

estimated and used to determine where to add it to any project list. Truly speaking, the approval of those above factors, development works can be launched.

□Operational Feasibility

□Economic Feasibility

□Technical Feasibility

Operational Feasibility

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

Economic Feasibility

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at anytime. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

Technical Feasibility

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The

system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and WebLogic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

2.4 SYSTEM DEVELOPMENT AND WORKING METHODOLOGIES

2.4.1 INPUT DESIGN

Info Design assumes an indispensable function in the existence pattern of programming improvement, it requires extremely cautious consideration of engineers. The information configuration is to take care of information to the application as precise as could be expected under the circumstances. So inputs should be planned viably so the blunders happening while at the same time taking care of are limited. As indicated by Software Engineering Concepts, the info structures or screens are intended to give to have an approval authority over as far as possible, extend and other related approvals.

This framework has input screens in practically all the modules. Blunder messages are created to alarm the client at whatever point he submits a few errors and aides him in the correct manner so invalid passages are not made. Let us see profoundly about this under module plan.

2.4.2 OUTPUT DESIGN

The Output from the PC is needed to essentially make an effective technique for correspondence inside the organization fundamentally among the venture chief

and his colleagues, as such, the head and the customers. The yield of VPN is the framework which permits the task head to deal with his customers as far as making new customers and doling out new undertakings to them, keeping up a record of the venture legitimacy and giving organizer level admittance to every customer on the client side contingent upon the activities designated to him. After finish of an undertaking, another task might be doled out to the customer. Client verification systems are kept up at the underlying stages itself. Another client might be made by the manager himself or a client would himself be able to enroll as another client yet the errand of appointing ventures and approving another client rests with the executive as it were.

The application begins running when it is executed just because. The worker must be begun and afterward the web voyager in utilized as the program. The task will run on the neighborhood so the worker machine will fill in as the manager while the other associated frameworks can go about as the customers. The created framework is profoundly easy to use and can be effortlessly perceived by anybody utilizing it in any event.

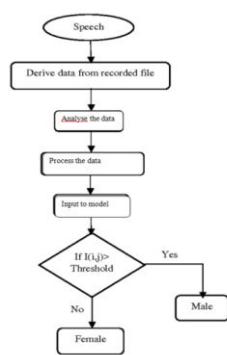


Fig. 1.4 FLOW CHART

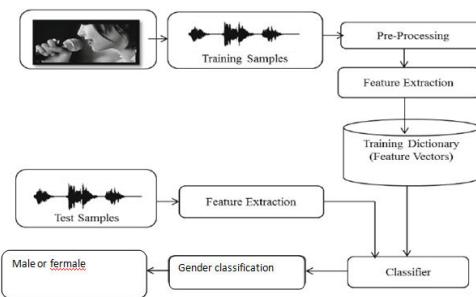


Fig.

1.5 BLOCK DIAGRAM

OUTPUT DESIGN

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining a record of the project validity and providing folder level access to each client on the user side depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself or a user can himself register as a new user but the task of assigning projects and validating a new user rests with the administrator only.

The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on

2.5 Proposed System

The system presents a differentially private sequential update of Bayesian

network (DPSUBN) approach for publishing horizontally partitioned high-dimensional data owned by multiple parties. We formally prove that DP-SUBN satisfies "-differential privacy. The idea of sequential update can also be adopted by other differentially private data processing tasks in the multi-party setting whenever the intermediate results obtained by some parties can be used as prior knowledge.

The system formulates the problem of search frontier construction as an optimization problem, and propose both exact and heuristic methods to solve this problem. In particular, for datasets with a moderate number of attributes, the exact method can obtain the optimal solution within reasonable time, while for datasets with a large number of attributes heuristic method can efficiently obtain an approximate solution. We show that both of these two methods are more effective than the search frontier construction method proposed in our previous work [9].

- The system develops a non-overlapping covering design (NOCD) method for generating all 2-way marginals of a given set of attributes. We further propose a dynamic programming method to find the optimal parameters used in NOCD. We show that NOCD achieves better data utility than the state-of-the-art technique.
- The system conducts an extensive experimental study over four real datasets. The experimental results demonstrate that DP-SUBN offers desirable data utility with low-communication-cost.

2.6 Advantages

- The system has more trust due to Bayesian Network and Search Frontier(DP-SUBN).
- The system calculates Less Communication cost in retrieving patients records

2.7 SYSTEM REQUIREMENTS

- **H/W System Configuration:-**
- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

2.8 SOFTWARE REQUIREMENTS

- | | |
|-------------------------|---------|
| Operating System | - |
| WindowsXP | |
| Coding Language | - |
| Java/J2EE(JSP, Servlet) | |
| Front End | - J2EE |
| Back End | - MySQL |

IMPLEMENTATION

4.1 ALGORITHMS

CONVOLUTIONAL NEURAL NETWORKS:

Convolutional Neural Network is one of the main categories to do image classification and image recognition in neural networks. Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolutional neural networks are widely used.

CNN takes an image as input, which is classified and process under a certain category such as dog, cat, lion, tiger, etc. The computer sees an image as an array of pixels and depends on the resolution of the image. Based on image resolution, it will see as $h * w * d$, where h = height w = width and d = dimension. For example, An RGB image is $6 * 6 * 3$ array of the matrix, and the grayscale image is $4 * 4 * 1$ array of the matrix.

In CNN, each input image will pass through a sequence of convolution layers along with pooling, fully connected layers, filters (Also known as kernels). After that, we will apply the Soft-max function to classify an object with probabilistic values 0 and 1.

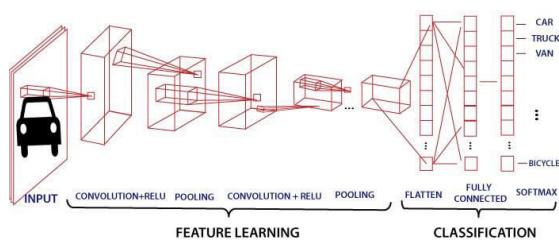


Fig. 2.3 CNN ARCHITECTURE

Convolution Layer

Convolution layer is the first layer to extract features from an input image. By learning image features using a small square of input data, the convolutional layer preserves the relationship between pixels. It is a mathematical operation which takes two inputs such as image matrix and a kernel or filter.

- The dimension of the image matrix is $h \times w \times d$.
- The dimension of the filter is $f_h \times f_w \times d$.
- The dimension of the output is $(h - f_h + 1) \times (w - f_w + 1) \times 1$.

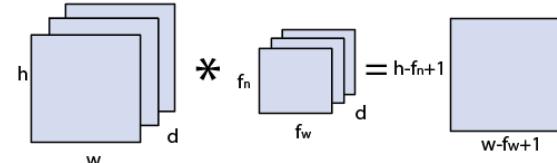


Image matrix multiplies kernel or filter matrix

Let's start with consideration a 5×5 image whose pixel values are 0, 1, and filter matrix 3×3 as:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

5×5 – Image Matrix 3×3 – Filter Matrix

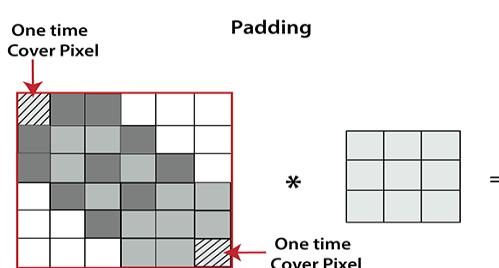
The convolution of 5×5 image matrix multiplies with 3×3 filter matrix is called "**Features Map**" and show as an output. Convolution of an image with different filters can perform an operation such as

blur, sharpen, and edge detection by applying filters.

Padding

Padding plays a crucial role in building the convolutional neural network. If the image will get shrink and if we will take a neural network with 100's of layers on it, it will give us a small image after filtered in the end.

If we take a three by three filter on top of a grayscale image and do the convolving then what will happen?



It is clear from the above picture that the pixel in the corner will only get covered one time, but the middle pixel will get covered more than once. It means that we have more information on that middle pixel, so there are two downsides:

- Shrinking outputs
- Losing information on the corner of the image.

To overcome this, we have introduced padding to an image. "**Padding is an additional layer which can add to the border of an image.**"

PoolingLayer

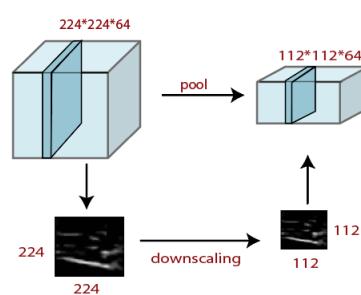
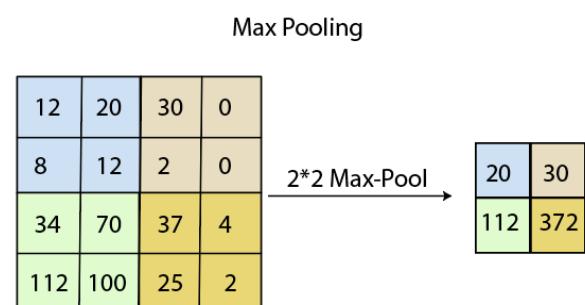
Pooling layer plays an important role in

pre-processing of an image. Pooling layer reduces the number of parameters when the images are too large. Pooling is "**downscaling**" of the image obtained from the previous layers. It can be compared to shrinking an image to reduce its pixel density. Spatial pooling is also called downsampling or subsampling, which reduces the dimensionality of each map but retains the important information. There are the following types of spatial pooling:

Max Pooling

Max pooling is a **sample-based discretization process**. Its main objective is to downscale an input representation, reducing its dimensionality and allowing for the assumption to be made about features contained in the sub-region binned.

Max pooling is done by applying a max filter to non-overlapping sub-regions of the initial representation.



Average Pooling

Down-scaling will perform through average pooling by dividing the input into rectangular pooling regions and computing the average values of each region.

Syntax

```
layer = averagePooling2dLayer(poolSize)
layer =
averagePooling2dLayer(poolSize,Name,Value)
Sum Pooling
```

The sub-region for **sum pooling** or **mean pooling** are set exactly the same as for **max-pooling** but instead of using the max function we use sum or mean.

Fully Connected Layer

The fully connected layer is a layer in which the input from the other layers will be flattened into a vector and sent. It will transform the output into the desired number of classes by the network.

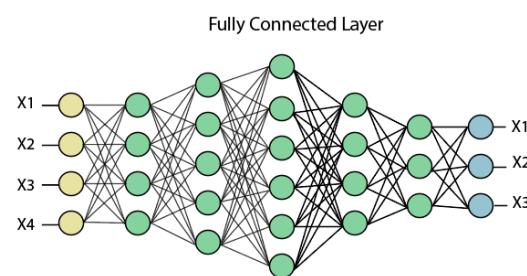


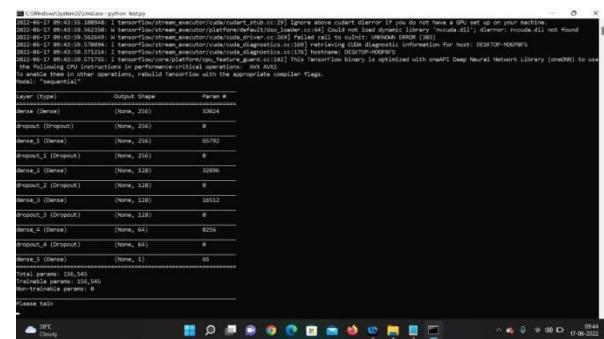
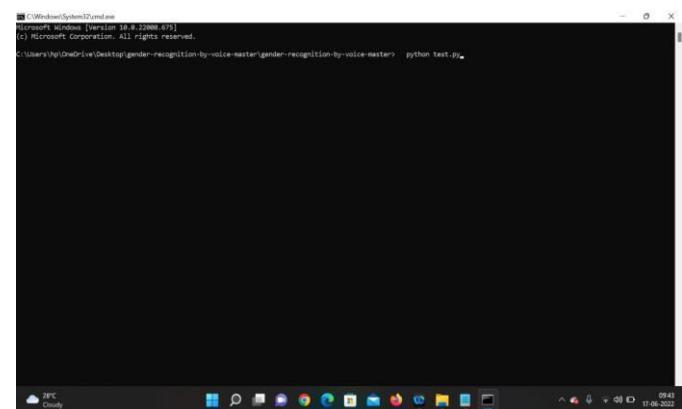
Fig. 2.4 FULLY CONNECTED LAYER

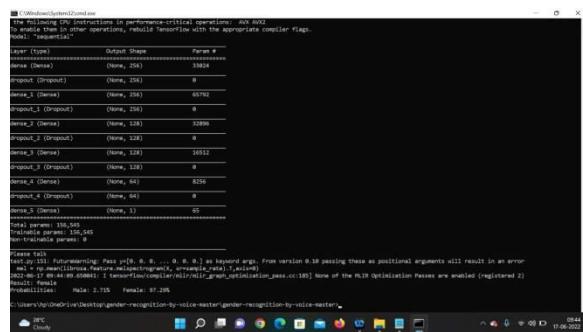
In the above diagram, the feature map matrix will be converted into the vector such as **x1**,

x2, **x3... xn** with the help of fully connected layers. We will combine features to create a model and apply the activation function such as **softmax** or **sigmoid** to classify the outputs as a car, dog, truck,etc.

RESULTS

Screenshots:





CONCLUSION AND FUTURE WORK

CONCLUSION AND FUTURE WORK:

In this work, we utilized a semi-supervised algorithm, called CNN for the gender recognition by voice. The proposed algorithm constitutes an ensemble of the most popular self-labeled algorithms i.e., Self-training, Co-training and Tri-training utilizing as base learner an ensemble of classifiers. The contribution of our approach as compared to other related approaches has to do with the fact that we utilize an ensemble of classifiers as base learner instead of single learners normally used in self-labeled algorithms. Our preliminary numerical results and the presented statistical analysis demonstrate the efficiency of the proposed algorithm for the gender recognition by voice compared against state-of-the-art self-labeled. Moreover, it presents competitive and sometimes better classification performance than classical supervised algorithms. Therefore, we conclude that reliable, stable and robust prediction models could be developed by the

adaptation of ensemble techniques in the semi-supervised learning framework and found that in our proposed approach there was no degradation in accuracy of prediction. Only an expected small increase in training time was noticed. Therefore, we can safely say that our proposed approach performs equally well even when input data increases drastically. Moreover, it is worth noticing that we understand the limitations imposed on the generalizability of the presented results due to the use of the only two freely available data. We certainly intend to investigate this further in the near future. Our future work is focused on improving the prediction accuracy of our framework by combining self-labeled algorithms with more sophisticated and theoretically motivated ensemble learning methodologies. Additionally, another interesting aspect is concentrating on extending our framework for handling big data with traditional technique and platforms such as Finally, since our numerical experiments are quite promising, we intend to focus on expanding our experiments and applying further the proposed algorithm to several audio datasets for speaker recognition

REFERENCES:

- [1] Pahwa, A.; Aggarwal, G. Speech feature extraction for gender recognition. *IJIGSP* 2016, 8, 17. [CrossRef]
- [2] Gamit, M.R.; Dhameliya, K.; Bhatt, N.S. Classification Techniques for Speech Recognition: A Review. *Int. J. Emerging Technol. Adv. Eng.* 2015, 5, 58–63.
- [3] Yasmin, G.; Dutta, S.; Ghosal, A. Discrimination of male and female

- voice using occurrence pattern of spectral flux. In Proceedings of the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 6–7 July 2017; pp. 576–581.
- [4] Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A.; Frà, C.; Valla, M. Spectra: A speech processing platform as smartphone application. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 7030–7035.
- [5] Wang, W.C.; Pestana, M.H.; Moutinho, L. The Effect of Emotions on Brand Recall by Gender Using Voice Emotion Response with Optimal Data Analysis. In Innovative Research Methodologies in Management; Springer: London, UK, 2018; pp. 103–133.
- [6] Holzinger, A. Introduction to machine learning and knowledge extraction (MAKE). *Mach. Learn. Knowl. Extr.* 2017, 1, 1–20. [CrossRef]
- [7] Ferri, M. Why topology for machine learning and knowledge extraction? *Mach. Learn. Knowl. Extr.* 2018, 1, 115–120. [CrossRef]
- [8] Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. Data Mining: Practical Machine Learning Tools and Techniques; Morgan Kaufmann: Burlington, MA, USA, 2016.
- [9] Buyukyilmaz, M.; Cibikdiken, A.O. Voice Gender Recognition Using Deep Learning. *Adv. Comput. Sci. Res.* 2016, 58, 409–411. Maka, T.; Dziurzanski, P. An analysis of the influence of acoustical adverse conditions on speaker gender identification. In Proceedings of the XXII Annual Pacific Voice Conference (PVC), Krakow, Poland, 11–13 April 2014; pp. 1–4.