

Find out Best machine learning algorithm to detect the presence of covid-19

Prashant Pawar , sonali Sethi

Computer Engineering, NBN SSoc

Submitted: 10-11-2021

Revised: 24-11-2021

Accepted: 27-11-2021

ABSTRACTION:No one had ever seen such an Pandemic. Millions of families took to the streets because of this covid19,finally our country knew the true state of our health system. This technology is very useful considering the huge stress on the health system. Corona is not gone yet,Corona wants to live with it so it can be used easily anywhere.

Using this ever-increasing artificial technology we can easily tell the difference between healthy and infected people Can recognize. We use different artificial intelligencetechnology to distinguish between Covid-19 patients and normal human beings and we take the required data from coswara data base.

Key words:Artificial intelligence techniques, Covid-19 detection, speech analysis, voice analysis.

Introduction: Covid-19 has devastated the world, millions of families have been devastated, the economy has collapsed, life has been short-lived. In this case, the patient needs to get the right service, as well as the difference between covid-19 and normal pneumonia, we can use the technology of ai better and faster, and in this we will study the same technique.

There is more than 240 Million cases noted in the world and more than 5 million deaths[1].

In recent times we have been using iot as well as ai in healthcare, personal care like fitness band [2] [3], also the day to day progress in this device depends on the progress of ai algorithm. Also the progress made in ai till date is health Remarkable in the field.[4]- [11].

We will use MACHINE LEARNING(ml) to find out whether the covid is positive or negative. Our main goal in this work will be to find the most accurate ml technic, using this technique to identify

whether the covid is covid at the initial stage, so that the treatment It will be easy to do. You can also use this technique with a smart phone.

IEEE Open Journal of A paper published in Engineering in Medicine and Biology states that people without covid symptoms are more at risk because they are covid patients even if they do not show symptoms. With the help of AI, we can distinguish between healthy and mainly asymptomatic people by recording cough. It has also recently been pointed out that the pronunciation of words is different when speaking between healthy and covid patients, mainly vowels.

MATERIAL METHODS:

1] In order to find the best ml tech, we are going to use the sound of coswara database specially vowel a, e, o, as well as we are also doing cough recording, additionally we are going to focus more on sound here. Indian Institute of Science (Bangalore) Realized coswara database in which sound records are collected from every continent except Africa.

The most important point of this database is that there is no language barrier, we can analyze the sounds and process them properly. [12]

We record all the records in the database at a frequency of 44.1 KHZ because the Nyquist – Shannon sampling theorem states that a sampling rate of more than twice the maximum frequency of the signal to be recorded is needed, resulting in a required rate of at at least 40 kHz. [13] usesopportunne filter to cancel outside noise.

In the table below you will see clear data which is used in the database:

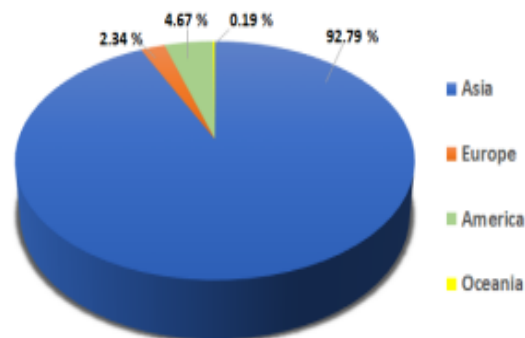


FIGURE 1. Geographical location of the subjects involved in this study.

2] We need to extract specific features from the sounds that we have recorded so that we can reach the desired result. In this we use different parameters like fundamental frequency, jitter, shimmer and Harmonic to Noise Ratio. [14], we use Mel-Frequency Cepstral Coefficients (MFCC) or Spectral Centroid or RollOff [15] - [18] to identify sound variations.

The fundamental frequency is nothing but reflects the vibrating rate of the vocal folds

Jitter & Shimmer means instability of the oscillating pattern of the vocal folds used to Evaluate the cycle-to-cycle changes in frequency (jitter) and amplitude (shimmer).

3] In this work we use different ml algorithm

Bayes theorem -for determining conditional probabilities and conditional probability to make accurate predictions in ml. Naive Bayes (NB) and BayesNet (BN) algorithms were used [19].

Support Vector Machine(SVM)-this uses for two group classification problem.

lazy -specific instances are used by these approaches to classify the data.

Meta -this algorithm combine multiple ml models for better classification result [20]

Rules-this technic divides data into some specific rule one r [21] or decision table [22]

Trees-used to classify categorical data

Result:

The performance of ml algorithm is calculated by the following term:

True positive: predicted value and actual value is positive.

True negative: predicted value and actual value is negative.

False positive: predicted value positive and actual value is negative.

False negative: predicted value negative and actual value is positive.

for results we uses different formulas:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}(1)}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2)$$

$$\text{F1-score} = \frac{2 * \text{precision} * \text{sensitivity}}{\text{precision} + \text{sensitivity}} \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4) \quad \text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

TABLE 2. Results achieved on the training set.

Category	Algorithm	Accuracy (%)	F1-score (%)	Specificity (%)	Precision (%)	Recall (%)	AUC
Bayes	NB	95.13	75.31	94.87	61.00	98.39	0.949
	BN	93.31	69.27	92.76	52.99	100	0.964
Functions	SVM	100	100	100	100	100	1.000
	SGD	100	100	100	100	100	1.000
Lazy	lbc	100	100	100	100	100	1.000
	LWL	93.80	70.86	93.29	54.87	100	0.999
Meta	Adaboost	98.18	89.21	98.03	80.52	100	0.998
	Bagging	98.66	91.20	99.21	90.48	91.94	0.998
Rules	OneR	96.47	74.78	98.68	81.13	69.35	0.840
	DT	97.93	85.95	99.08	88.14	83.87	0.981
Tree	J48	99.39	95.93	99.74	96.72	95.16	0.998
	REPTree	95.99	74.81	97.37	71.01	79.03	0.972

TABLE 3. Results achieved on the testing set.

Category	Algorithm	Accuracy (%)	F1-score (%)	Specificity (%)	Precision (%)	Recall (%)	AUC
Bayes	NB	90.24	54.55	91.05	41.38	80.00	0.861
	BN	89.76	58.82	88.95	41.67	100	0.945
Functions	SVM	97.07	82.35	97.37	73.68	93.33	0.954
	SGD	95.12	66.67	97.37	66.67	66.67	0.820
Lazy	lbc	93.17	50.00	96.84	53.85	46.67	0.718
	LWL	89.76	58.82	88.95	41.67	100	0.971
Meta	Adaboost	91.22	55.00	92.63	44.00	73.33	0.963
	Bagging	92.20	50.00	95.26	47.06	53.33	0.965
Rules	OneR	88.78	34.29	92.63	30.00	40.00	0.663
	DT	91.22	35.71	95.79	38.46	33.33	0.829
Tree	J48	93.66	60.61	95.79	55.56	66.67	0.954
	REPTree	91.22	52.63	93.16	43.48	66.67	0.948

CONCLUSION:

The health system was in turmoil due to Covid-19, which spread like wildfire all over the world, people literally died without any bed or

oxygen. The search for a vaccine is on all over the world, but we do not yet have a 100% effective vaccine. At such times, we can test Covid-19 and

diagnose and treat it properly, without any contact with anyone, at a very fast speed.

The increasing use of iot and ai is revolutionizing every field every day, the future is ai and iotchach so we can use it to treat many more diseases.

Technology has played an important role in this fast paced world of man.

As part of that, you can use ml, what technique, algorithm for Covid-19 is accurate and easy for you, also using this technology, it is easy to test your Covid-19 for very little money (almost free).

Our main objective was to find out which algorithm gives the best results using all these algorithms, without concluding that the svm algorithm is the most accurate which is up to 97% accurate. It gives us output accordingly, so we can test our Covid-19 by believing in it.

REFERENCES:

- [1]. World Health Organization. (2020) WHO Coronavirus Disease (COVID-19) Dashboard. Accessed: Jan. 1, 2021. [Online]. Available: https://covid19.who.int/?gclid=AlaIQobChMIht_qyL_K6gIVB7tCh2AlgwMEAAAYASAAEgLyX_D_BwE/
- [2]. G. Fortino, S. Galzarano, R. Gravina, and W. Li, "A framework for collaborative computing and multi-sensor data fusion in body sensor networks," *Inf. Fusion*, vol. 22, pp. 50–70, Mar. 2015
- [3]. R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.
- [4]. M. S. Hossain, "Cloud-supported cyber-physical localization framework for patients monitoring," *IEEE Syst. J.*, vol. 11, no. 1, pp. 118–127, Mar. 2017.
- [5]. L. Verde, G. De Pietro, P. Veltri, and G. Sannino, "An m-health system for the estimation of voice disorders," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops (ICMEW)*, Jun. 2015, pp. 1–6.
- [6]. M. S. Hossain and G. Muhammad, "Emotion-aware connected healthcare big data towards 5G," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2399–2406, Aug. 2018.
- [7]. U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "Voice disorder detection via an M-health system: Design and results of a clinical study to evaluate Vox4Health," *BioMed Res. Int.*, vol. 2018, pp. 1–19, Aug. 2018.
- [8]. G. Muhammad, M. S. Hossain, and N. Kumar, "EEG-based pathology detection for home health monitoring," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 603–610, Feb. 2021.
- [9]. M. S. Hossain and G. Muhammad, "Deep learning based pathology detection for smart connected healthcare," *IEEE Netw.*, vol. 34, no. 6, pp. 120–125, Nov. 2020.
- [10]. A. Yassine, S. Singh, M. S. Hossain, and G. Muhammad, "IoT big data analytics for smart homes with fog and cloud computing," *Future Gener. Comput. Syst.*, vol. 91, pp. 563–573, Feb. 2019.
- [11]. A. Amini, W. Chen, G. Fortino, Y. Li, Y. Pan, and M. D. Wang, "Editorial special issue on 'AI-driven informatics, sensing, imaging and big data analytics for fighting the COVID-19 pandemic,'" *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2731–2732, Oct. 2020.
- [12]. A. Tsanas, "Acoustic analysis toolkit for biomedical speech signal processing: Concepts and algorithms," in *Models and Analysis of Vocal Emissions for Biomedical Applications*, vol. 2. Firenze, Italy: Firenze Univ. Press, 2013, pp. 37–40.
- [13]. https://en.wikipedia.org/wiki/44,100_Hz
- [14]. A. R. Maccarini and E. Lucchini, "La valutazione soggettiva ed oggettiva della disfonìa. Il protocollo sifel," *Acta Phoniatria Latina*, vol. 24, nos. 1–2, pp. 13–42, 2002.
- [15]. L. Verde, G. De Pietro, and G. Sannino, "Voice disorder identification by using machine learning techniques," *IEEE Access*, vol. 6, pp. 16246–16255, 2018.
- [16]. M. Pishgar, F. Karim, S. Majumdar, and H. Darabi, "Pathological voice classification using mel-cepstrum vectors and support vector machine," 2018, arXiv:1812.07729. [Online]. Available: <http://arxiv.org/abs/1812.07729>
- [17]. J. C. Saldanha, T. Ananthakrishna, and R. Pinto, "Vocal fold pathology assessment using mel-frequency cepstral coefficients and linear predictive cepstral coefficients features," *J. Med. Imag. health Informat.*, vol. 4, no. 2, pp. 168–173, 2014.
- [18]. V. Gupta, "Voice disorder detection using long short term memory (LSTM) model," 2018, arXiv:1812.01779. [Online]. Available: <http://arxiv.org/abs/1812.01779>
- [19]. G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in *Proc. Conf. Uncertainty*

- Artif. Intell. San Mateo, CA, USA: Morgan Kaufmann, 1995, pp. 338–345.
- [20]. T. G. Dietterich, “Ensemble methods in machine learning,” in Proc. Int. workshop multiple classifier Syst. Cagliari, Italy: Springer, 2000, pp. 1–15.
- [21]. R. Holte, “Very simple classification rules perform well on most commonly used datasets,” Mach. Learn., vol. 11, pp. 63–91, 1993.
- [22]. R. Kohavi, “The power of decision tables,” in Proc. 8th Eur. Conf. Mach. Learn. Crete, Greece: Springer, 1995, pp. 174–189.
- [23]. <https://ieeexplore.ieee.org/document/941646>
- 9