

# Penggabungan Metode Itakura Saito Distance dan Backpropagation Neural Network untuk Peningkatan Akurasi Suara pada Audio Forensik

## *(Combining Itakura Saito Distance and Backpropagation Neural Network Methods to Improve Sound Accuracy in Audio Forensic)*

Ardy Wicaksono<sup>1</sup>, Sisdarmanto Adinandra<sup>2</sup>, Yudi Prayudi<sup>3</sup>

<sup>1,2,3</sup>Program Studi Magister Informatika, Universitas Islam Indonesia, Jogjakarta

<sup>1</sup>ardy.wicaksono@students.uui.ac.id

<sup>2</sup>s.adinandra@uui.ac.id

<sup>3</sup>prayudi@uui.ac.id

**Abstrak**—Audio merupakan salah satu barang bukti digital yang digunakan dalam *cybercrime*. Seringkali bahwa bukti *audio* ini membawa peran krusial untuk mengungkapkan adanya kasus kejahatan sehingga diperlukan proses analisis *audio forensic*. *Audio* tersebut berisi rekaman suara seseorang yang memiliki karakter dengan pengucapan kosakata yang berbeda-beda, pengucapan yang tidak jelas, dan memiliki banyak *noise*. Perlu adanya penanganan yang sesuai *Standart Operational Procedure (SOP) audio forensics*. Tahapan dalam melakukan *audio forensic* pada *Digital Forensic Analyst Team* Pusat Laboratorium Forensik (DFAT) PUSLABFOR terdiri dari 4 tahapan yakni *Acquisition, Audio Enhancement, Decoding, dan Voice Recognition*. Dalam penelitian ini akan dibahas mengenai analisis *audio* menggunakan metode *speech processing* yaitu *Itakura Saito Distance* dan metode jaringan syaraf tiruan yaitu *Backpropagation Neural Network* dengan tujuan memperkuat hasil akurasi identik suatu barang bukti rekaman suara. Jika metode ini dikaloraborasi akan memperkuat tingkat akurasi dan *argument* yang diperoleh dari proses analisa, khususnya dalam penanganan *audio forensic*. Akurasi itu sendiri diukur dari nilai kedekatan frekuensi atau *spectrum* antara rekaman suara asli dengan rekaman suara pembanding. Hasil pengujian yang dilakukan pada 4 rekaman suara asli (*unknown*) dan 4 rekaman suara pembanding (*known*) dengan lebih dari 20 kosakata menunjukkan akurasi tertinggi yang identik lebih dari 95%.

**Kata-kata kunci:** *itakura saito distance, backpropagation neural network, audio forensik, akurasi suara*

**Abstract**—Audio is one of the digital evidence used in *cybercrime*. Often, this audio evidence plays a crucial role in revealing a crime case, so an audio forensics analysis process is required. The audio contains a recorded voice of a person who has a character with different vocabulary pronounced, unclear pronunciation, and has a lot of noise. There needs to be an appropriate handling of the Audio forensics Standard Operational Procedure (SOP). The stages in conducting audio forensic on the Digital Forensic Analyst Team of the Forensic Laboratory Center (DFAT) PUSLABFOR consist of 4 stages, namely Acquisition, Audio Enhancement, Decoding, and Voice Recognition. This research will discuss about audio analysis using speech processing method, namely Itakura Saito Distance and artificial neural network method, namely Backpropagation Neural Network with the aim of strengthening the identical accuracy results of a sound recording evidence. If this method is used, it will strengthen the level of accuracy and arguments obtained from the analysis process, especially in handling audio forensics. Accuracy itself is measured from the frequency proximity value or spectrum between the original sound recording and the comparator sound recording. The results of tests carried out on 4 unknown voice recordings and 4 known voice recordings with more than 20 vocabularies show the highest identical accuracy of more than 95%.

**Keyword:** *itakura saito distance, backpropagation neural network, audio forensic, sound accuracy*

## I. PENDAHULUAN

Audio merupakan suara yang dihasilkan oleh getaran benda atau suatu benda yang berupa sinyal analog dengan *amplitude* yang berubah secara terus menerus terhadap

waktu dan berkaitan dengan indra pendengaran yang merambat melalui udara [1]. *Audio* yang berupa rekaman suara memiliki karakter pengucapan kosakata yang berbeda-beda dan kondisi yang tidak jelas. Rekaman suara berisi gelombang bunyi yang direkam dengan teknologi *digital recording* [2]. Suara diakibatkan dari suatu getaran yang menghasilkan pola gelombang tertentu, karena perubahan tekanan udara secara cepat [3]. Suara mempunyai komponen yaitu *pitch*, *formant*, dan *spectrogram* yang dapat digunakan untuk mengidentifikasi karakteristik suara seseorang untuk kepentingan *voice recognition* [4].

*Audio* sering dijadikan salah satu jenis barang bukti digital. Barang bukti yang diajukan dipersidangan harus terjamin untuk mencegah gangguan bukti, khususnya barang bukti digital [5]. Demi memenuhi keuntungan beberapa orang dipengadilan rekaman percakapan dari barang bukti *audio* sering sekali dilakukan manipulasi atau pengeditan [6]. Dalam beberapa kasus, *audio* menjadi barang bukti yang sering menjadi perdebatan khususnya dalam mengukur dan mendeteksi keakuratan suara. Sehingga dengan adanya hal tersebut diperlukan pendukung keputusan dalam memverifikasi *audio forensic* [7]. Kedudukan barang bukti digital sangat penting dalam pembuktian suatu kasus, sehingga diperlukan saksi ahli yang kompeten khususnya dibidang keilmuan *audio forensic* [8]. Seorang saksi ahli harus mampu memaparkan analisis yang relevan menyeluruh dan otentik dalam melakukan serangkaian pemeriksaan. Membuktikan suatu peristiwa khususnya barang bukti *audio* harus sesuai dengan ketentuan, karena menyangkut nasib dari korban maupun tersangka [9]. Menggunakan teknik *audio forensic* layak untuk digunakan dalam menganalisis rekaman suara untuk menentukan kepemilikan suara [10]. Serangkaian pemeriksaan harus mengikuti Undang-Undang, SOP, dan sesuai dengan ketentuan *audio forensic* pada DFAT PUSLABFOR. Analisis yang harus dilakukan pada *audio forensic* yaitu *acquisition*, *audio enhancement*, *decoding*, dan *voice recognition*, dimana Pengucapan kata-kata yang akan dianalisis minimal mendapatkan 20 kata untuk dapat menunjukkan keidentikan suara *unknown* (asli) dengan *known* (pembanding) [4].

Kemampuan untuk melakukan analisis *audio forensic* tergantung pada ketersediaan rekaman *audio* yang diperoleh. Rekaman *audio* berisi wawancara, penyadapan rahasia, serta merekam interogasi [11]. Rekaman *audio* yang ditemukan memiliki karakteristik berbeda-beda pada setiap orangnya dan rekaman yang diperoleh kualitasnya belum tentu baik. Rekaman suara dengan merendahkan *pitch* cenderung mudah diidentifikasi, dibandingkan rekaman meninggikan *pitch* [12]. Jika terjadi kasus tindakan kriminal dan ditemukan barang bukti *audio* berupa rekaman suara yang terduga sebagai pelaku, maka perlu adanya proses analisis untuk mencari

kebenaran rekaman suara yang ditemukan identik dengan rekaman suara pelaku. Analisis barang bukti digital yang berupa *audio* dapat dilakukan dengan metode berbasis *autorecorder* atau *automatic statistic* yaitu *Itakura Saito Distance*. Metode ini menggunakan jarak antara *spectrum* suara asli dan suara perkiraan sebagai indikator akurasi barang bukti *audio* [13]. *Autoencoder* pada suara yang bergema merekonstruksi pengenalan ucapan dengan cara *spectrum* jangka pendek dan jangka panjang diuji secara matematis, sehingga memperoleh akurasi dari rekaman suara yang diperoleh [14].

Metode *Itakura Saito Distance* merupakan metode *speech processing* yang dapat menggunakan *linier prediction coding* dalam mengukur perbedaan *spectrum* suara asli dan *spectrum* suara perkiraan. Metode ini cukup memegang peranan dalam hal kajian mengenai *speech coding*, *analysis*, *synthesis* and *recognition*. *Itakura Saito* dikenal juga dengan *maximum likelihood distortion distance*, dimana *linier prediction* (LP) dijadikan perkiraan kemungkinan. *Linier prediction* yaitu bagaimana mengoptimalkan kecepatan *bit* yang rendah dengan kualitas suara memadai [15].  $d_{1S}$  adalah variabel dari *Itakura Saito* yang dapat dilihat pada persamaan  $d_{1S}(S, S') = \int_{-\pi}^{\pi} \frac{S(w)}{S'(w)} \frac{dw}{2\pi} - \log \frac{\sigma_{\infty}^2}{\sigma_{\infty'}^2} - 1$ .

Dimana  $\sigma_{\infty}^2$  *spectrum* sinyal pertama dari rekaman suara *unknown* dan  $\sigma_{\infty'}^2$  *spectrum* sinyal suara perkiraan dari rekaman suara *known* yang merupakan *prediction errors* pada  $S(w)$   $S'(w)$ . *Maximum likelihood* dari *Itakura Saito* merupakan langkah-langkah distorsi ucapan pada *spectrum* suara asli dan perkiraan [16]. Dari proses perhitungan LPC dengan metode *Itakura Saito Distance* dapat diperoleh nilai akurasi yang identik apabila semakin dekat jarak antara dua *spectrum*, maka semakin mirip perbandingan suara tersebut [17].

Perolehan hasil akurasi tidak lepas dari analisa *anova*. Analisa *anova* adalah analisa yang mengkalkulasi secara statistik nilai-nilai *formant* dari suara *unknown* dan suara *known*. Kalkulasi nilai *formant* menggunakan *Likelihood Ratio* dapat dilihat pada persamaan  $LR = \frac{p(E|H_p)}{p(E|H_d)}$ .

Dimana  $p(E|H_p)$  berasal dari *p-value anova* merupakan hipotesis tuntutan (*prosecution*), yaitu sampel suara *known* dan *unknown* berasal dari orang sama. Sedangkan  $p(E|H_d) = 1 - p(E|H_p)$  merupakan hipotesis perlawanan, yaitu suara *known* dan *unknown* berasal dari orang yang berbeda [4]. Nilai presentase tinggi akurasi dari metode *Itakura Saito Distance* masih menjadi perdebatan. Asumsi tersebut disampaikan oleh seorang saksi ahli dalam kasus tindak pidana korupsi di Indonesia tentang barang bukti rekaman suara, dimana metode ini banyak digunakan dalam rekayasa *audio* yaitu *background* musik ataupun konser, tetapi tidak

mengarah untuk membandingkan dua suara [18]. Analisis yang sudah dilakukan dari proses identifikasi rekaman suara asli dengan rekaman suara pembandingan dengan metode *Itakura Saito Distance* dan LPC memperoleh nilai presentase akurasi identik tertinggi sebesar 90% [17].

Metode *Backpropagation Neural Network* (BPNN) merupakan algoritma yang efektif dalam memberikan solusi penurunan gradien untuk meminimalkan kuadrat *error input*. Metode BPNN mempunyai tiga tahapan utama yaitu tahap perambatan maju, tahap perambatan balik, dan tahap perubahan bobot dan bias [19]. *Backpropagation* sering dikatakan algoritma *multilayer* yang mempunyai parameter yaitu *input layer*, *hidden layer*, dan *output layer*. Pada *input layer* tidak terjadi proses perhitungan, tetapi terjadi pengiriman sinyal *input* ke  $x$  *hidden layer*. *Hidden layer* pada *backpropagation* dapat menyebabkan tingkat *error* menjadi lebih kecil. Hal itu disebabkan sebagai tempat untuk memperbaharui dan menyesuaikan bobot yang diinginkan. Bias diarahkan mendekati dengan target *output* yang diinginkan [20]. Gradien dari fungsi dihitung sehubungan dengan bobot dan dalam bias jaringan yang meminimalkan fungsi tersebut [13]. Tujuan utama metode BPNN adalah mendapatkan keseimbangan antara pengenalan pola pelatihan secara benar dan respon yang baik untuk pola lain yang sejenis. Umumnya pola data dibagi menjadi dua bagian, yaitu pola data yang dipakai sebagai pelatihan dan data yang dipakai untuk pengujian. Perubahan bobot dilakukan berdasarkan pola pelatihan. Akan tetapi selama pelatihan dalam *epoch* tertentu, kesalahan yang terjadi dihitung berdasarkan data pelatihan dan pengujian [21]. Jika nilai *error* menurun maka akan melakukan proses pelatihan selanjutnya [22]. Akan tetapi jika kesalahannya sudah meningkat, pelatihan tidak diteruskan lagi. Ini sejalan dengan hasil penelitian [23] yang menyatakan bahwa algoritma pelatihan memiliki kinerja terbaik saat menggunakan *learning rate*=0.5.

Kasus tindakan kriminal yang melibatkan barang bukti rekaman suara masih menjadi perdebatan tersendiri dari bentuk karakter suara yang dihasilkan dari rekaman *audio*, maupun algoritma yang belum pasti mengenai perbandingan identik atau tidak rekaman suara tersebut. Metode yang sudah disebutkan diatas memiliki karakteristik dan parameter yang berbeda-beda, akan tetapi memiliki tujuan yang sama. Mengasumsikan metode *Itakura Saito Distance* yang mempunyai nilai identik yang lebih baik antara perbandingan suara, kemudian hasil dari metode tersebut diekstraksi kembali dengan metode BPNN, sehingga akan mendapatkan nilai yang relevan dalam menganalisa bukti digital *file* rekaman *audio* sebagai barang bukti digital sesuai dengan ketentuan *Chain of Custody* (CoC). Perolehan akurasi hasil identik antara dua suara dari metode *Itakura Saito Distance* masih ada kemungkinan ditingkatkan kembali

nilai akurasi dengan menggunakan metode *Backpropagation Neural Network*, sekaligus memperkuat *argument* penanganan dalam proses *audio forensic*.

## II. METODE

### A. Pengambilan sampel data

Sesuai dengan SOP tentang analisis *audio forensic* pengambilan dilakukan dengan alat perekam suara dengan pengucapan minimal 20 kata. Hal itu agar dapat menunjukkan keidentikan suara asli dengan suara pembandingan. Pengambilan sampel suara menghasilkan rekaman suara dalam format *.wav*. Rekaman suara tersebut selanjutnya dijadikan parameter pada setiap metode yang akan dianalisis. Proses analisis menggunakan skema penanganan barang bukti rekaman suara pada *audio forensic*. Tahapan penanganan barang bukti *audio forensic* yang diperoleh berdasarkan skenario kasus dengan memiliki beberapa tahapan sesuai dengan SOP yaitu:

1) *Aquisition*: proses akuisisi dilakukan terhadap barang bukti elektronik yang berisikan rekaman suara asli (*unknown*) yang ditemukan ditempat kejadian perkara maupun barang bukti rekaman yang akan sebagai rekaman suara pembandingan (*known*). Berdasarkan skenario rekaman suara yang ditemukan dari proses akuisisi *audio recorder* akan menghasilkan *file DD image* sesuai dengan yang ditetapkan, dengan penemuan *file* yang berekstensi *.wav*.

2) *Audio Enhancement*: tahapan ini merupakan proses evaluasi rekaman suara yang ditemukan dengan memperdengarkan (*playback*) untuk mengetahui kualitas suaranya. Jika rekaman suara kualitas suara kurang baik mungkin dikarenakan adanya faktor *noise*, maka diperlukan proses untuk menghilangkan *noise* tersebut. Penelitian ini menggunakan algoritma *lowpass filter* untuk menstabilkan *amplifier* dengan menghilangkan *high frequency* pada rekaman suara, sekaligus menggunakan *tools* Adobe Audition CC 2019.

3) *Decoding*: tahap ini adalah proses transkrip dari rekaman suara yang ditemukan, dan pembuatan transkrip ini dilakukan dengan tujuan membantu mendapatkan nilai akurasi yang lebih presisi terhadap hasil transkrip. Penelitian ini menggunakan proses transkrip secara online melalui situs [otranscribe.com](http://otranscribe.com).

4) *Voice Recognition*: *voice Recognition* adalah teknik dimana proses analisis atau identifikasi apakah rekaman suara asli (*Unknown*) identik dengan rekaman suara pembandingan (*Known*) yang didapatkan dari tahapan *sampling*. Analisis dilakukan menggunakan Matlab 2015b. Sesuai dengan ketentuan *audio forensic*,

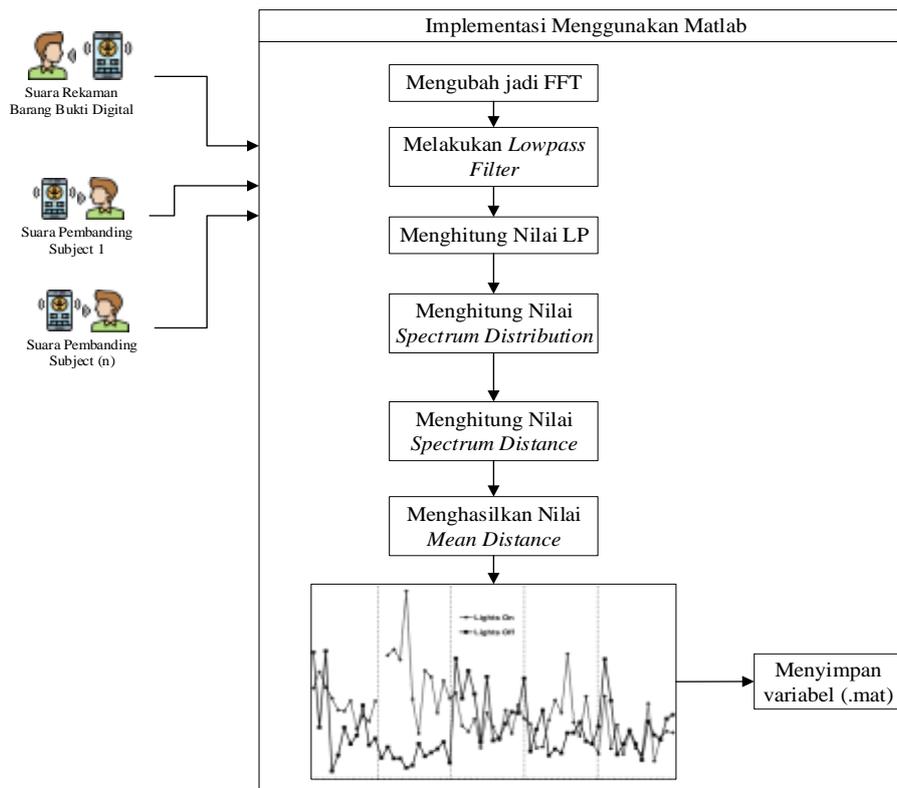
pada proses ini harus memiliki minimal 20 kata yang memiliki kesamaan antara rekaman suara barang bukti (*unknown*) dengan rekaman suara pembanding (*known*). Penelitian ini akan melakukan analisis rekaman suara pada proses *voice recognition* yaitu menggunakan Metode *Itakura-saito Distance* dan Metode *Backpropagation Neural Network*.

**B. Implementasi Metode**

Metode *Itakura Saito Distance* memiliki beberapa tahapan dalam proses analisis yaitu hasil rekaman suara dari proses *audio enhancement* dengan format *.wav*. Rekaman suara menjadi *input* berupa sinyal analog. Sinyal analog diubah menjadi sinyal frekuensi (*spectrum*) dengan FFT. Selanjutnya melakukan tahap *lowpass filter* untuk menghilangkan *high frequency noise* yang masih terdapat pada *audio enhancement* [24]. Tahap selanjutnya mengevaluasi kualitas bicara berdasarkan *linear prediction* antara rekaman suara (A1) dan rekaman suara pembanding (P1) yang dianalisis dengan melakukan perhitungan LPC. Berikut skema proses implementasi metode *Itakura Saito Distance* dapat dilihat pada Gambar 1.

Untuk menghitung LPC dibutuhkan rentang waktu tertentu (*frame*) dan menghitung jarak *spectrum* antara

dua rekaman suara. Selanjutnya akan diperoleh nilai rentan jarak untuk menentukan rekaman suara identik atau tidak, kemudian nilai yang diperoleh menjadi ekstraksi ciri (nilai target) dan dianalisa dengan metode *Backpropagation Neural Network*. Algoritma *backpropagation* menggunakan fungsi *sigmoid biner*, karena *output* yang diharapkan bernilai 0 sampai 1. Lamanya interasi yang harus dilakukan menjadi masalah utama dalam algoritma *Backpropagation*. *Backpropagation* tidak dapat memberikan kepastian berapa nilai *epoch* yang harus dilalui untuk mencapai kondisi diinginkan, sehingga parameter yang ada dari setiap peneliti berbeda-beda dalam menghasilkan jumlah iterasi yang relatif lebih sedikit. Perintah yang dipakai dalam membentuk jaringan dapat dilihat pada persamaan  $= newff(PR, [S1 \dots Sn], \{Tf1 \dots Tfn\}, Btf, Blf, Pf)$ . Dimana *net* merupakan jaringan *n layer*. *PR* merupakan matriks ordo  $R \times 2$ . *Sn* merupakan jumlah unit *n layer* (*tansig*). *Tfn* merupakan fungsi aktivasi yang dipakai *n layer* (*traingdx*). *Btf* merupakan fungsi pelatihan jaringan (*learnrgrm*). *Blf* merupakan fungsi perubahan bobot/bias. *Pf* merupakan fungsi perhitungan *error* (MSE).



**Gambar 1.** Tahap implementasi metode *Itakura Saito Distance*

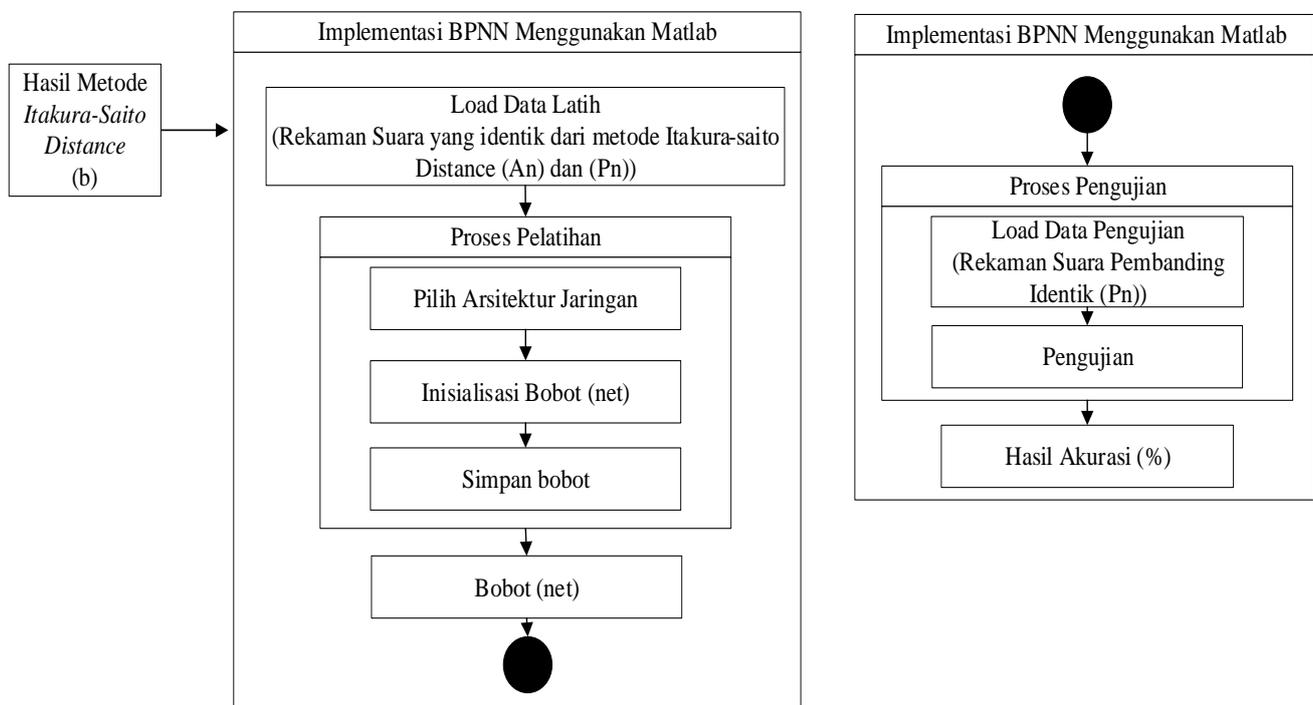
Pemilihan bobot awal sangat mempengaruhi terhadap nilai *error*, serta cepat tidaknya proses pelatihan. Apabila nilai bobot awal terlalu besar, maka *input* ke setiap *hidden layer* dan *output layer* akan jatuh pada daerah fungsi *sigmoid* yang sangat kecil. Sebaliknya, apabila nilai bobot awal terlalu kecil, maka *hidden layer* dan *output layer* akan sangat kecil, sehingga menyebabkan proses pelatihan berjalan lambat. Bobot awal diinisialisasi secara *random* dengan nilai antara -0.5 sampai 0.5 interval lainnya. Penentuan jumlah *hidden layer* membuat pelatihan *Backpropagation* menjadi lebih mudah. Dalam perambatan maju, keluaran harus dihitung untuk tiap *layer*, dimulai dari *hidden layer* paling bawah (terdekat dengan *input*). Dalam perampatan mundur, faktor  $\delta$  perlu dihitung untuk tiap *hidden layer*, dimulai dari *output layer*. Skema implementasi metode *Backpropagation Neural Network* dapat dilihat pada Gambar 2. Metode tersebut diharapkan dapat memperoleh hasil perbandingan suara menjadi akurat. Pengecekan *error* dengan cara menghitung nilai *Mean Square Error* (MSE), apabila  $MSE_{input} \leq MSE$  maka proses dihentikan, jika tidak maka proses akan diteruskan untuk membaca data pertama lagi ( $i = 1$ ) hingga sarat  $MSE_{hitung} \leq MSE_{input}$  terpenuhi. Selanjutnya pada tahap pengujian, dimana akan dilakukan perhitungan semua *hidden layer* dan *output layer*.

### III. HASIL DAN PEMBAHASAN

#### A. Data penelitian

Implementasi metode yang ada diperlukan skenario kasus secara eksperimen dengan melibatkan rekaman suara asli (*unknown*) dan rekaman suara pembanding (*known*). Eksperimen skenario kasus diperoleh sebuah barang bukti rekaman suara yang akan dianalisis. Melalui tahapan *decoding* diperoleh transkrip secara lengkap isi percakapan rekaman suara yang terduga sebagai tersangka. Selanjutnya transkrip rekaman dianalisis untuk menentukan bagian percakapan dengan kata-kata lebih dari 20 kata.

Berikut sampel rekaman suara yang digunakan pada penelitian ini yaitu “Kalau dapat bukan berarti tinggal disitu, ya kalau sudah ada warga negara itu kan kamu lepas warga negara Indonesia. Nah, ini berarti kan kamu tidak bisa di buru lagi, gitu loh, selesai. Kamu bisa pergi ke seluruh dunia, mau di mana-mana saja”. Setelah menentukan kalimat panjang yang akan dijadikan analisis, selanjutnya mencari rekaman suara dari beberapa pembanding. Dimana untuk memperkuat asumsi dari implementasi metode yang digunakan, penelitian ini menggunakan 4 rekaman suara *known* dan rekaman suara *unknown*. Implementasi metode dilakukan pada tahapan *voice recognition* dengan menggunakan aplikasi Matlab 2015b yaitu:

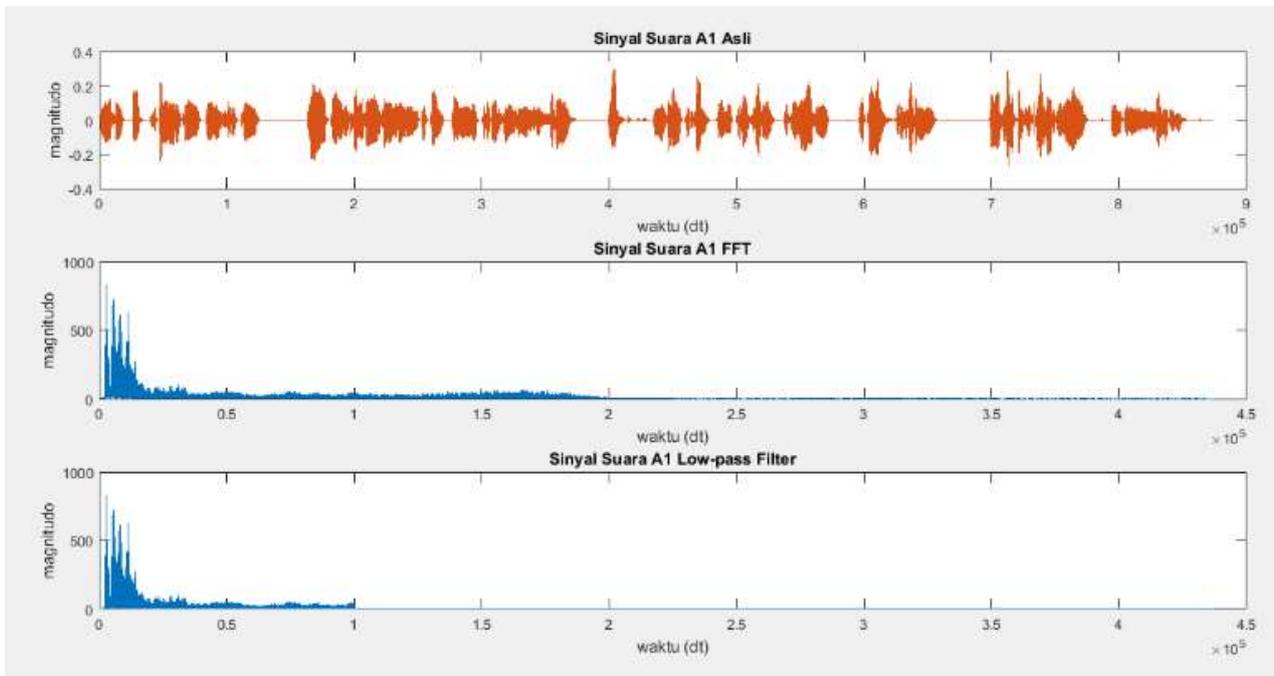


Gambar 2. Tahap implementasi metode *backpropagation neural network*

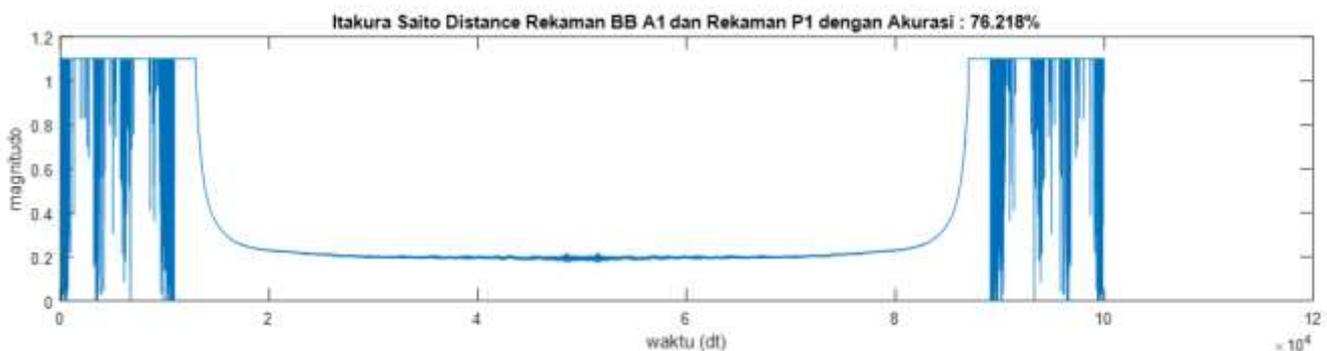
**B. Metode Itakura Saito Distance**

*Itakura Saito Distance* mencerminkan perbedaan *spectrum amplitude* waktu antara jarak dari dua rekaman suara, sehingga dapat mendistorsi sinyal ucapan dengan baik. Rekaman suara yang sudah dilakukan tahap *audio enhancement* akan menjadi *input* berupa sinyal analog. Sinyal analog tersebut diubah kedalam bentuk sinyal frekuensi menggunakan algoritma FFT. Untuk memperbaiki kualitas suara yang masih terdapat pada *audio enhancement* dilakukan proses *Lowpass Filter*. Penelitian ini menggunakan nilai batas panjang waktu (*frame*) sebesar 100000 dari jumlah keseluruhan panjang waktu dari setiap suara itu sendiri. Berikut hasil dari sampel satu suara “A1” yang dilakukan proses FFT dan *Lowpass Filter* dapat dilihat pada Gambar 3.

Pada Gambar 3, *Lowpass Filter* terlihat perbedaan pengurangan frekuensi menjadi 0 dari rentang waktu 1 ke seterusnya. selanjutnya mengevaluasi kualitas bicara berdasarkan *linear prediction* antara rekaman suara (A1) dan rekaman suara pembanding (P1) yang dianalisis dengan melakukan perhitungan dengan *function* LPC. Hasil kalkulasi nilai LPC kemudian diproses dengan beberapa *function* seperti *lpcar2ra*, *lpcar2rr*, *lpcar2rf*, *distisar* untuk menghitung jarak dua *spectrum*. Selanjutnya akan diperoleh nilai kedekatan frekuensi atau *spectrum* antara rekaman suara A1 dan P1 yang disebut nilai akurasi. Dua rekaman suara tersebut memiliki tingkat akurasi sebesar 76,218% dengan 100000 *frame* dari jumlah keseluruhan *frame*. Hasil tersebut dapat dilihat dari visualisasi yang terdapat pada Gambar 4.



**Gambar 3. Hasil dari memperbaiki kualitas suara dengan low-pass filter**



**Gambar 4. Hasil Visualisasi Itakura-Saito Distance**

Berikut keseluruhan hasil dari empat rekaman suara asli (*unknown*) dari empat responden dan empat rekaman suara pembandingan (*known*) dari empat responden, dapat dilihat pada

TABEL I.

C. Metode *backpropagation neural network*

Berdasarkan hasil dari metode *Itakura Saito Distance* selanjutnya dianalisis kembali untuk mendapatkan tingkat akurasi yang lebih tinggi lagi. Sebagai contoh suara A1 dan P3 identik yang disimpan dengan variabel b3 dengan format ekstensi *mat*. Rekaman suara yang identik dari masing-masing perbandingan akan jadi acuan untuk meningkatkan akurasi. Berikut proses dari metode *Backpropagation Neural Network* yaitu:

1) *Proses training Backpropagation Neural Network*

Proses pertama dalam metode *Backpropagation Neural Network* yaitu *load data* dua rekaman suara yang diduga identik dan akan dijadikan data *training*. Sama halnya dengan proses yang ada di metode *Itakura Saito Distance*, data yang akan diuji harus dinormalisasi terlebih dahulu dengan mengubah sinyal ke domain frekuensi (FFT) dan memperbaiki kualitas suara memakai *low-pass filter*. Kemudian diperoleh hasil nilai  $x1$  dan  $y1$  dilakukan *invers* matriks yang nilai ini akan dijadikan nilai *input*. Tahap selanjutnya adalah memasukkan parameter untuk proses melakukan *training*. Selanjutnya proses *training* akan berjalan seperti yang terlihat pada Gambar 5.

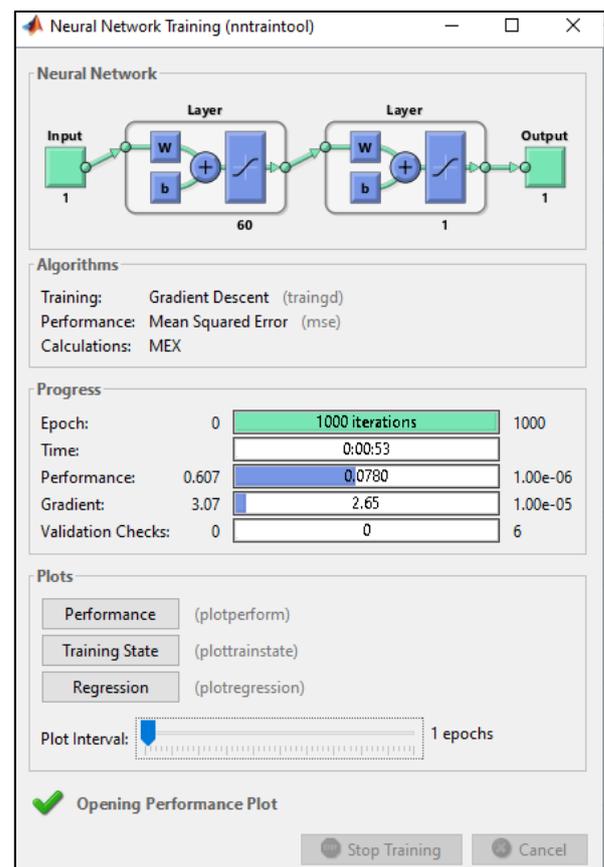
Terlihat beberapa informasi proses keseluruhan dengan jumlah 100000 *input* membutuhkan waktu 53 detik dengan *epoch* (pelatihan) 1000, MSE 0,000001 dan batas *gradient* 0,00001 diperoleh *performance* sebesar 0,0780. Sehingga diperoleh diperoleh detail informasi, dimana *best training performance* sebesar 0,077985 pada *epoch* 995.

2) *Proses pengujian Backpropagation Neural Network*

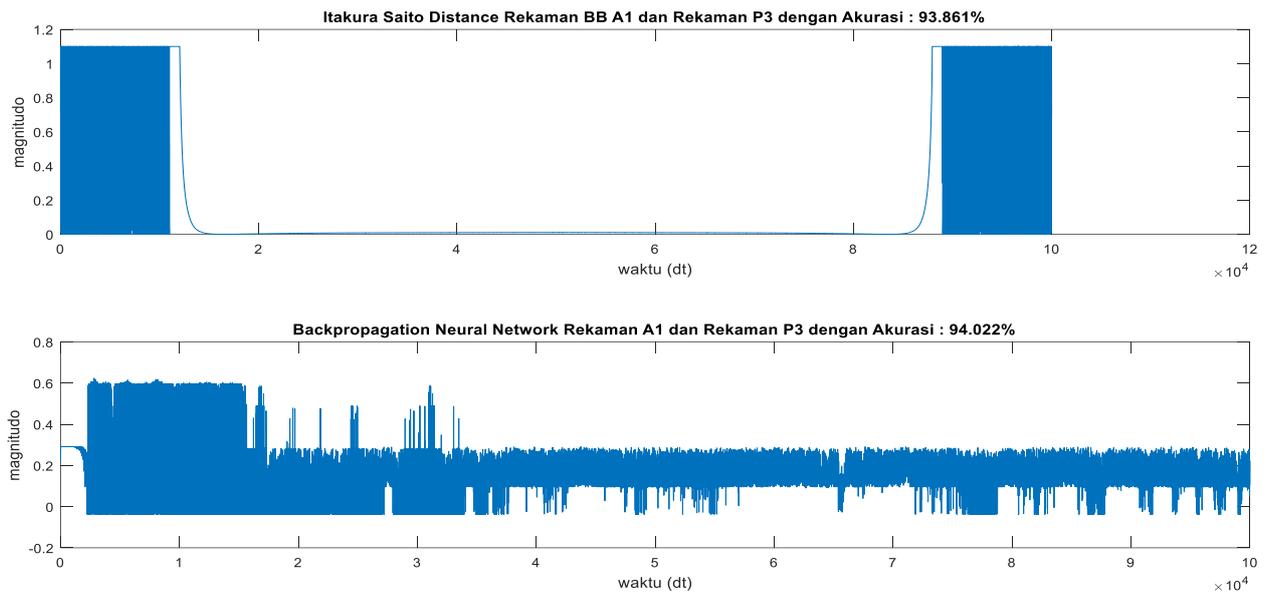
Setelah melakukan *training* selanjutnya adalah tahap pengujian dari rekaman suara pembandingan identik dari proses metode *Itakura Saito Distance* akan jadikan *load data*. Seperti tahap sebelumnya pertama melakukan perubahan sinyal domain frekuensi dan melakukan *Lowpass filter*. Dimana nilai *training* dengan variabel “*net*” dilakukan proses pengujian P3. Berikut hasil dari metode *Backpropagation Neural Network* pada rekaman suara P3 sekaligus dibandingkan langsung dengan hasil metode *Itakura-Saito Distance* dapat dilihat pada Gambar 6.

TABEL I  
HASIL EVALUASI PERBANDINGAN NILAI ITAKURA-SAITO DISTANCE PADA REKAMAN *UNKNOWN* DAN REKAMAN *KNOWN*

Subjek Rekaman Suara	Presentase Akurasi	Keterangan
A1	P1 76,218%	Rekaman A1 Identik dengan Rekaman P3
	P2 76,423%	
	P3 93,861%	
	P4 77,317%	
A2	P1 76,667%	Rekaman A2 Identik dengan Rekaman P2
	P2 89,312%	
	P3 68,535%	
	P4 87,459%	
A3	P1 84,805%	Rekaman A3 Identik dengan Rekaman P4
	P2 92,812%	
	P3 73,030%	
	P4 95,952%	
A4	P1 94,011%	Rekaman A4 Identik dengan Rekaman P1
	P2 85,724%	
	P3 83,763%	
	P4 87,639%	



Gambar 5. Proses *training Backpropagation Neural Network*



Gambar 6. Hasil visualisasi

Rekaman suara A1 dengan rekaman suara P3 dengan metode *itakura-saito distance* sudah dikatakan identik dengan nilai sebesar 93,861%, sedangkan dengan menggunakan metode *backpropagation neural network* mengalami peningkatan akurasi sebesar 94,022%.

D. Hasil Analisis Metode Itakura Saito Distance dengan Metode Backpropagation Neural Network

Implementasi metode *Itakura Saito Distance* dengan metode *Backpropagation Neural Network* dalam mencari nilai akurasi identik rekaman suara *known* dan rekaman suara *unknown* dapat dilihat pada TABEL II.

Dilihat dari TABEL II rata-rata rekaman suara *unknown* ( $A_n$ ) dan *known* ( $P_n$ ) yang dikatakan identik mengalami peningkatan akurasi. Hal ini sesuai dengan penelitian yang dilakukan sebelumnya, dimana analisa rekaman suara *unknown* dan *known* dengan menggunakan metode *Itakura Saito Distance* diperoleh presentase hasil akurasi tertinggi lebih dari 90% [17]. Proses analisa perbandingan suara *unknown* dan *known* diharapkan dapat memperkuat hasil akurasi dan *argument* dalam penanganan *audio forensic*. Sehingga metode *Itakura Saito Distance* didukung dengan metode *Backpropagation Neural Network* dapat diterima dalam persidangan yang melibatkan barang bukti rekaman suara.

TABEL II  
HASIL AKURASI REKAMAN SUARA DENGAN METODE *ITAKURA-SAITO DISTANCE* DAN *BACKPROPAGATION NEURAL NETWORK*

Subjek Suara	Ket		Akurasi (%)	
			Metode <i>Itakura-Saito Distance</i> FFT <i>Low-Pass Filter</i> LPC	Metode <i>Backpropagation Neural Network</i> FFT <i>Low-Pass Filter</i> -
A1	P3	Identik	93,861%	94,022%
A2	P2	Identik	89,312%	90,939%
A3	P4	Identik	95,952%	95,117%
A4	P1	Identik	94,011%	94,432%

IV. PENUTUP

Metode *Itakura-Saito Distance* sangat dipengaruhi panjang waktu dan kualitas dari rekaman suara tersebut, sehingga memakan waktu yang cukup lama dalam proses analisisnya. Sedangkan untuk metode *Backpropagation Neural Network* dengan kondisi yang sama, cukup efektif dalam menganalisis rekaman suara, akan tetapi dalam proses analisisnya sangat dipengaruhi ketepatan penentuan parameter. Kedua metode tersebut memiliki persamaan yaitu untuk mendapatkan hasil yang sesuai harus menggunakan metode lain seperti *Fast Faurier Transform* untuk mengubah sinyal analog ke domain frekuensi dan *Lowpass Filter* untuk peningkatan kualitas rekaman suara. Menggabungkan metode *speech*

processing menggunakan Itakura Saito Distance dan metode jaringan syaraf tiruan dengan metode Backpropagation Neural Network akan mendapatkan nilai akurasi yang lebih efisien atau tinggi dalam mencari rekaman suara yang identik antara suara asli dan suara pembanding, sekaligus memperkuat argument dalam penanganan audio forensic. Hasil nilai akurasi suara yang identik memperoleh presentase tertinggi 95%. Peningkatan tersebut disebabkan dari faktor analisis parameter dari proses training yang tepat.

#### DAFTAR PUSTAKA

- [1] A. Imario, D. W. Sudiharto, and E. Ariyanto, "Uji Validasi Suara Berbasis Pengenalan Suara (Voice Recognition) Menggunakan Easy VR 3.0," *Pros. SNATIF ke-4 Tahun 2017*, pp. 153–160, 2017.
- [2] Mansyur, "Alat Bukti Rekaman Suara dalam Pembuktian Perkara Tindak Pidana Korupsi," *J. Komun. Huk.*, vol. 3, no. 1, 2017.
- [3] B. W. Prasetya, B. Susanto, and J. Purwadi, "Identifikasi Suara Pria dan Wanita Berdasarkan Frekuensi Suara," *J. Inform.*, vol. 4, no. 1, pp. 2–9, 2011.
- [4] M. Al-Azhar Nuh, "Audio Forensic: Theory and Analysis," pp. 1–38, 2011.
- [5] Y. Prayudi, A. Luthfi, and A. M. R. Pratama, "Pendekatan Model Ontologi Untuk Merepresentasikan Body of Knowledge Digital Chain of Custody," *Cybermatika ITB*, vol. 2, no. 2, pp. 36–43, 2014.
- [6] M. Imran, Z. Ali, S. T. Bakhsh, and S. Akram, "Blind Detection of Copy-Move Forgery in Digital Audio Forensics," *IEEE Access*, vol. 5, pp. 12843–12855, 2017.
- [7] D. Renza, D. M. Ballesteros L., and C. Lemus, "Authenticity verification of audio signals based on fragile watermarking for audio forensics," *Expert Syst. Appl.*, vol. 91, pp. 211–222, Jan. 2018.
- [8] C. Handoko, "Kedudukan Alat Bukti Digital Dalam Pembuktian Cybercrime Di Pengadilan," *J. Jurisprud.*, vol. 6, no. 1, p. 1, 2017.
- [9] Andri and D. A. D. Tawang, "Validitas Keterangan Ahli dari Penyidik dalam Pembuktian Kasus Ujaran Kebencian Ras DI Media Sosial (Studi Putusan: NO.1105/PID.SUS/2017/PN JKT.UTR)," *J. Huk. Adigama*, vol. 1, 2018.
- [10] V. R. C. Putri and Sunarno, "Analisis Rekaman Suara Menggunakan Teknik Audio Forensik Untuk Keperluan Barang Bukti Digital," *Unnes Phys. J.*, vol. 3, no. 1, 2014.
- [11] R. C. Maher, "History of Audio Forensics .Principles of Forensic Audio Analysis," pp. 29–37, 2018.
- [12] A. Subki, B. Sugiantoro, and Y. Prayudi, "Analisis Rekaman Suara Voice Changer dan Rekaman Suara Asli Menggunakan Metode Audio Forensik," *Indones. J. Netw. Secur.*, vol. 7, no. 1, 2018.
- [13] Y. Nozaki and T. Nakamoto, "Itakura-Saito distance based autoencoder for dimensionality reduction of mass spectra," *Chemom. Intell. Lab. Syst.*, vol. 167, pp. 63–68, Aug. 2017.
- [14] T. Ishii, H. Komiyama, T. Shinozaki, Y. Horiuchi, and S. Kuroiwa, "Reverberant Speech Recognition Based on Denoising Autoencoder Division of Information Sciences , Graduate School of Advanced Integration Science ," *Interspeech*, no. August, pp. 3512–3516, 2013.
- [15] R. M. Gray, "Packet speech on the Arpanet: A history of early LPC speech and its accidental impact on the Internet Protocol," *Electr. Eng.*, no. June, pp. 1–44, 2007.
- [16] B. K. Padhy and S. K. Sahu, "Analysis of Speech Recognition Techniques," 2009.
- [17] A. Irawan, "Perbandingan Metode Itakura-Saito Distance dan Manual Statistik (Pitch, Formant, Spectrogram) untuk Akurasi Identifikasi Suara pada Audio Forensik," Universitas Islam Indonesia, 2019.
- [18] Hukumonline.com, "Metode 'Itakura-Saito', Silang Pendapat Dua Ahli Forensik di Sidang Tipikor," 2019. [Online]. Available: <https://www.hukumonline.com/berita/baca/lt5c5ced64ac1a4/metode-itakura-saito--silang-pendapat-dua-ahli-forensik-di-sidang-tipikor?page=3>. [Accessed: 13-Jun-2020].
- [19] H.-X. Huang, J.-C. Li, and C.-L. Xiao, "A proposed iteration optimization approach integrating backpropagation neural network with genetic algorithm," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 146–155, Jan. 2015.
- [20] M. N. Sidqi, "Sistem Cerdas Deteksi Suara Untuk Pengklasifikasian Smart Sound Detection System For Classifying Heart Disease," *J. Elektron. Pendidik. Tek. Elektron.*, vol. 8, no. 1, pp. 42–55, 2018.
- [21] M. A. Hilmi, "Identifikasi Suara Menggunakan Fft Dan Neural Network," Universitas Muhammadiyah Gresik, 2014.
- [22] T. Sutojo, E. Mulyanto, and V. Suhartono, *Kecerdasan Buatan*. Yogyakarta: Andi Offset, 2011.
- [23] H. Mustafidah and S. Suwarsito, "Performance of Levenberg-Marquardt Algorithm in Backpropagation Network Based on the Number of Neurons in Hidden Layers and Learning Rate," *JUITA J. Inform.*, vol. 8, no. 1, p. 29, 2020.
- [24] M. Singh and E. N. Kumar Garg, "Audio Noise Reduction Using Butter Worth Filter," *Int. J. Comput. Organ. Trends*, vol. 6, no. 1, pp. 20–23, 2014.

