

## Deteksi Onset Gamelan Bebas DWPT dan BLSTM

Hisyam Mustofa\*<sup>1</sup>, Agfianto Eko Putra<sup>2</sup>

<sup>1</sup>Magister Ilmu Komputer, UGM, Yogyakarta, Indonesia

<sup>2</sup>Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

e-mail: \*<sup>1</sup>mustofahisyam13@mail.ugm.ac.id, <sup>2</sup>agfi@ugm.ac.id

### Abstrak

Gamelan terdiri dari berbagai macam instrumen yang memiliki karakteristik yang berbeda. Masing-masing memiliki karakteristik dalam hal frekuensi dasar amplitudo, sinyal envelope, serta perbedaan cara memainkannya menghasilkan perbedaan daya sustain sinyal. Karakteristik tersebut menimbulkan masalah vanishing gradient pada model Elman Network yang digunakan pada penelitian sebelumnya dalam mempelajari pendeteksian onset yang ada dalam sinyal instrumen Saron yang memiliki rata-rata interval lebih dari 0,6 detik. Penelitian ini menggunakan BLSTM (Bidirectional Long Short Term Memory) sebagai model untuk pelatihan dan Transformasi Wavelet Packet untuk mendesain psychoacoustic critical bandwidth sebagai model untuk ekstraksi fitur. Untuk metode peak picking penelitian ini menggunakan metode ambang batas fix dengan nilai 0.25. Penggunaan model BLSTM yang didukung oleh Transformasi Wavelet Packet diharapkan dapat mengatasi vanishing gradient yang ada pada arsitektur RNN sederhana. Model diuji berdasarkan 3 parameter evaluasi yaitu precision, recall dan F-Measure. Berdasarkan skenario pengujian yang dilakukan, model dapat mengatasi permasalahan vanishing gradient pada instrumen saron yang memiliki rata-rata interval antar onset 600 ms. Dari total onset sebanyak 428 pada instrumen saron, model berhasil mendeteksi 426 secara benar, dengan 4 onset yang salah terdeteksi dan 2 onset yang tidak terdeteksi. Evaluasi menyeluruh untuk masing-masing precision, recall, dan F1-Measure algoritma ini mendapatkan 0,975, 0,945 dan 0,960.

**Kata kunci**— Deteksi Onset, Transformasi Wavelet Packet, Bidirectional Long Short Term Memory.

### Abstract

Gamelan consists of various kinds of instruments that have different characteristics. Each has characteristics in terms of the basic frequency, amplitude, signal envelope, and different ways of playing it, resulting in differences in the sustain power of the signal. These characteristics cause the problem of vanishing gradient in the Elman Network model which was used in previous studies in studying the onset detection in the Saron instrument signal which has an average interval of more than 0.6 seconds. This study uses BLSTM (Bidirectional Long Short Term Memory) as a model for training and Wavelet Packet Transformation to design a psychoacoustic critical bandwidth as a model for feature extraction. For the peak picking method, this study uses a fixed threshold method with a value of 0.25. The use of the BLSTM model supported by the Wavelet Packet Transform is expected to overcome the vanishing gradient that exists in a simple RNN architecture. The model was tested based on 3 evaluation parameters, namely precision, recall and F-Measure. Based on the test scenario carried out, the model can overcome the vanishing gradient problem on the Saron instrument which has an average interval between onset of 600 ms. Out of a total of 428 onsets on the Saron instrument, the model successfully detected 426 correctly, with 4 incorrectly detected onsets and 2 undetected onsets. A thorough evaluation for each of the precision, recall, and F1-Measure algorithms obtained 0.975, 0.945 and 0.960.

**Keywords**— Onset Detection, Wavelet Packet Transformation, Bidirectional Long Short Term Memory.

## 1. PENDAHULUAN

Dalam perspektif musik, onset secara sederhana dapat didefinisikan sebagai titik awal sebuah nada. Onset memiliki peranan penting dalam perspektif warna musik (*timbre*). Hal tersebut membuat pendeteksian onset secara otomatis berkembang dalam berbagai bidang penelitian seperti pendengaran mesin, pemrosesan music dan *music information retrieval* [1]. Menurut Stasiak onset detection merupakan atribut penting yang dimiliki oleh sistem intelegensi sebagai tahapan fundamental dasar untuk analisis audio, ekstraksi fitur, deskripsi, sehingga akhirnya memberikan suatu informasi yang komprehensif [2].

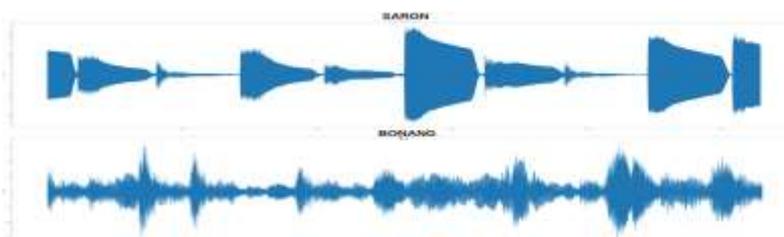
Deteksi musik onset memiliki kegunaan yang luas untuk membangun berbagai macam aplikasi seperti analisi tempo dan *beat tracking*, transkripsi music otomatis [3], identifikasi instrument, pergantian harmonic, dan generator metadata untuk membangun system rekomendasi khusus music pada mesin pencarian [1]. Karena kegunaan tersebut, penelitian tentang deteksi onset menarik minat banyak peneliti dari berbagai bidang.

Gamelan adalah jenis kelompok musik terdiri dari berbagai macam instrument seperti saron, demung, peking, dan slenthem, bonang, kenong, kempul, gong, dan sebagainya. Masing-masing memiliki karakteristik dalam hal frekuensi dasar amplitudo, sinyal *envelope*, serta perbedaan cara memainkannya menghasilkan perbedaan daya sustain sinyal. Selain itu gamelan tidak memiliki standar penyetelan nada, sehingga setiap antar daerah dapat memiliki frekuensi yang berbeda pada nada yang sama[4]. Karakteristik tersebut memberikan tantangan tersendiri dalam pendeteksian onset secara otomatis.

Pada Gambar 1 ditunjukkan contoh sinyal dari instrumen Saron dan Bonang. Karakteristik instrumen Saron memiliki interval onset (not) yang cukup jauh dibanding dengan instrument bonang tetapi memiliki daya sustain yang pendek. Sedangkan instrument Bonang memiliki interval onset yang cukup pender dengan daya sustain sinyal yang panjang. Kedua hal tersebut disebabkan tipe instrument dan cara memainkannya.

Penelitian terkait deteksi onset menggunakan metode *supervised learning* menggunakan model *Recurrent Neural Network* [5]. Penelitian tersebut berfokus pada perbandingan performat BTRL dan BPTT dalam melatih *Recurrent Neural Network*. Kekurangan pada penelitian tersebut adalah informasi yang ada didalamnya hanya sebatas proses training sehingga tidak ada informasi mengenai model tersebut dievaluasi menggunakan data tes.

Penelitian ini menggunakan *supervised learning* dengan metode yang berbeda dari penelitian sebelumnya. Data sinyal diekstrak fiturnya menggunakan pendekatan *psychoacoustics critical bandwidth* yang didesain menggunakan Transformasi Paket Wavelet. Fitur tersebut kemudian dilatih menggunakan model BLSTM (*Bidirectional Long Short Term Memory*). Model tersebut kemudian dievaluasi menggunakan 3 parameter yaitu *precision*, *recall*, dan *F-Measure*.



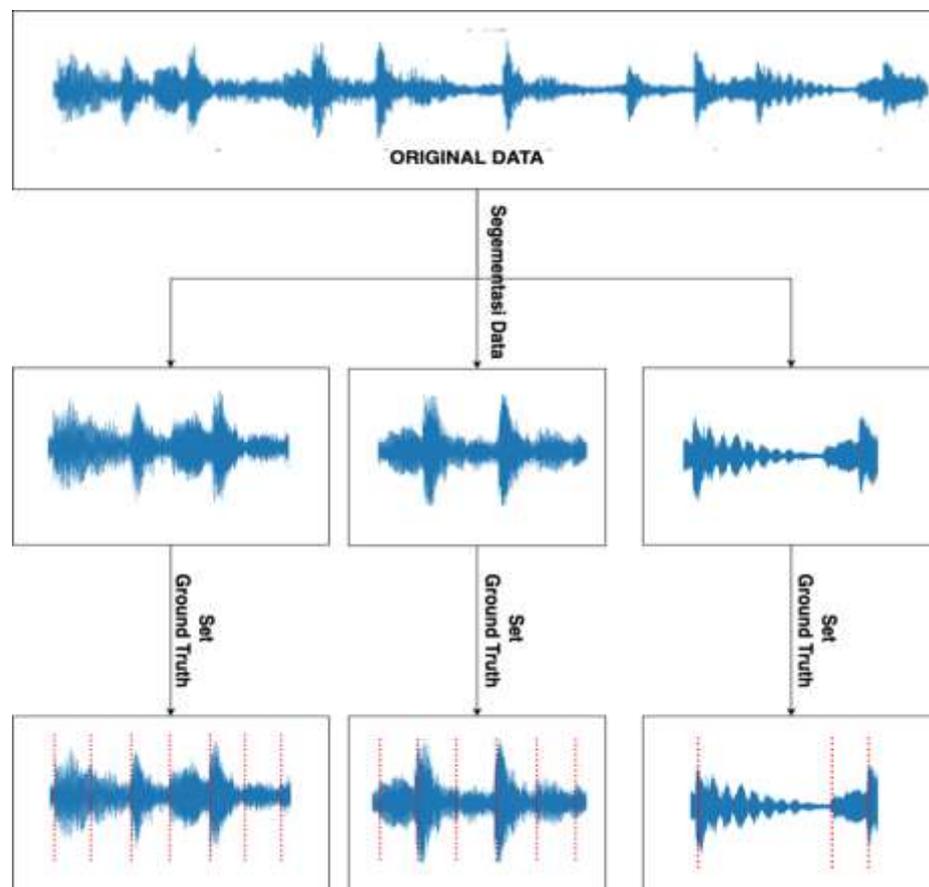
Gambar 1 Sinyal Instrumen Saron dan Bonang

## 2. METODE PENELITIAN

Deteksi onset yang dilakukan dalam penelitian menggunakan model algoritma pembelajaran BLSTM (*Bidirectional Long Short Term Memory*). Secara umum dibagi menjadi 4 tahapan yaitu 1) Pembuatan dataset dan *ground truth*. 2) *Preprocessing* (Ekstraksi Fitur) 3) Pelatihan. 4) Pengujian.

### 2.1 Pembuatan Ground Truth (Dataset)

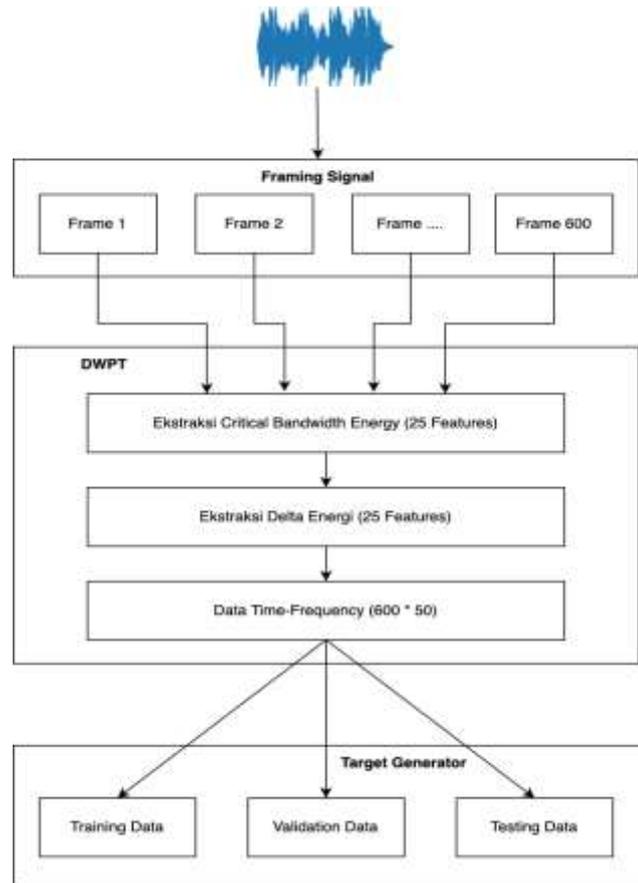
Proses pembuatan dataset dimulai dari pengumpulan data berupa sinyal audio yang direkam. Karena proses pengumpulan data tidak berada pada studio yang kedap suara, selanjutnya dilakukan proses *filtering* supaya hasil datanya lebih optimal. Proses *filter* sinyal berperan untuk membuang frekuensi-frekuensi yang tidak dibutuhkan pada proses analisis. Setelah proses *filtering* selesai, dilakukan proses segmentasi data: Setiap data akan dibatasi durasinya masing-masing 6 detik. Karena durasi data yang direkam lebih dari 6 menit, data perlu disegmentasi ke dalam beberapa bagian. Proses segmentasi data dilakukan secara manual dibantu dengan *script*. Beberapa bagian dengan kualitas yang baik atau rendah *noise* akan dipilih. Langkah terakhir adalah proses pelabelan data atau pembuatan *ground truth*. Pembuatan *ground truth* dilakukan secara manual dengan dibantu aplikasi pengolah audio. *Ground truth* perlu dibuat secara teliti karena berperan terhadap performa yang dibuat. Pembuatan *ground truth* akan menghasilkan data-data waktu kapan onset terjadi. Proses pembuatan dataset ditunjukkan pada Gambar 2.



Gambar 2 Pembuatan *Ground Truth* dan Dataset

### 2.2 Preprocessing (Ekstaksi Fitur)

Sebelum masuk ke algoritma pembelajaran BLSTM (*Bidirectional Long Short Term Memory*) masing masing audio dilakukan preprocessing menggunakan transformasi wavelet packet. Pendekatan ini menghitung sub band energi pada setiap *frame* dengan mengekstrak komponen approximation dan detail [6]. Tahapan pada proses *preprocessing* ditunjukkan pada Gambar 2.

Gambar 2 Proses *Preprocessing*

### 2.2.1 Framing Sinyal

Sebelum proses ekstraksi perlu dilakukan pembagian sinyal ke dalam *frame*. Proses ini bertujuan untuk merestrukturisasi data dengan melokalisasi data dalam jangka waktu-waktu tertentu sehingga dapat dilakukan proses analisis *time-frequency*. Proses ini menggunakan metode Hanning window dengan ukuran jendela 2048 yang umum digunakan dalam proses deteksi onset [7]. Pertama-tama sinyal masukan suara  $x[n]$  disegmentasikan ke dalam bingkai  $W = 2.048$  sampel yang sesuai dengan 46ms. Fungsi standar *windowing* kemudian diterapkan pada setiap *frame* seperti dengan memilih *frame rate*  $F_f = 100\text{fps}$ , ukuran hop  $h$  antara jendela yang berdekatan sama dengan  $\frac{F_s}{F_f}$  di mana  $F_s$  menunjukkan *sample rate* (yaitu  $F_s = 44,1 \text{ kHz}$ ) dan mereka tumpang tindih dengan faktor  $(W h)/W$ . Proses ini akan menghasilkan 600 frame sebagai representasi data waktu. Sample tersebut kemudian akan diekstraksi fitur-fiturnya yaitu frekuensi dan energinya menggunakan DWPT.

### 2.2.2 Ekstraksi Fitur (*Discrete Wavelet Packet Transform*)

Ekstraksi fitur pada penelitian ini menggunakan *Discrete Wavelet Packet Transform* (DWPT). DWPT adalah generalisasi dari *discrete wavelet transform* yang mentransformasi sinyal dari domain waktu ke dalam domain frekuensi-waktu. Berbeda dengan transformasi wavelet yang hanya membagi subband frekuensi rendah (*approximation coefficient*), *wavelet packet transform* juga membagi subband frekuensi tinggi (*details coefficient*).

Desain dari DWPT mengacu pada karakter pendengaran manusia terhadap frekuensi, setiap masukan dari *frame* dari suatu sinyal akan dihitung energi sub-bands nya dalam domain wavelet sesuai dengan desain *psychoacoustics critical bandwidth*. Pembagian sub-bands ditunjukkan pada Tabel 1.

Tabel 1 Pembagian sub bands berdasarkan *psychoacoustic critical bandwidths*

Dec. Level	Level Bandwidth	N. Bands	Resolusi Frekuensi
8	0 ÷ 516 Hz	6	86,13 Hz
7	516 ÷ 1378 Hz	5	172,27 Hz
6	1.38 ÷ 3.45 kHz	6	344,53 Hz
5	3.45 ÷ 5.51 kHz	3	689,06 Hz
4	5.51 ÷ 8.27 kHz	2	1378,13 Hz
3	8.27 ÷ 11 kHz	1	2756,25 Hz
2	11 ÷ 22 kHz	2	5512,50 Hz
Total	0 ÷ 22 kHz	25	

Pembagian pada Tabel 1 menghasilkan fitur sebanyak 25 sesuai dengan pembagian Total N. Bands. Energi pada masing-masing bands kemudian diekstrak dan di normalisasi menggunakan fungsi logaritmik untuk meniru karakter sensitivitas pendengaran manusia sesuai dengan persamaan 1 dimana  $n$  adalah index frame dan  $l$  adalah index bandwidth. Koefisien delta tambahan yang merujuk pada *Wavelet Packet Energy Coefficient* (WPEC) ditambahkan untuk mengekstraksi energi yang ada pada setiap sub bands sesuai dengan persamaan 2. Ekstraksi ini menghasilkan 25 fitur tambahan sehingga total fitur per frame-nya adalah 50. Setelah proses ini selesai, fitur yang dihasilkan adalah  $600 \times 50$ .

$$WPEC'(n, l) = \log(E(n, l) + 1.0) \quad (1)$$

$$WPEC''(n, l) = WPEC'(n, l) - WPEC'(n - 2, l) \quad (2)$$

### 2.2.3 Data Fold

*Data fold* membagi data ke dalam 3 kategori yaitu data latih (*training*), data validasi (*validation*) dan data pengujian (*test*). Pembagian data menggunakan skema *k-fold cross validation* dengan jumlah *fold* 6 dengan proporsi data latih, validasi dan pengujian yaitu 4:1:1 [8]. Skema *data fold* pada penelitian ini ditunjukkan pada Gambar 3.

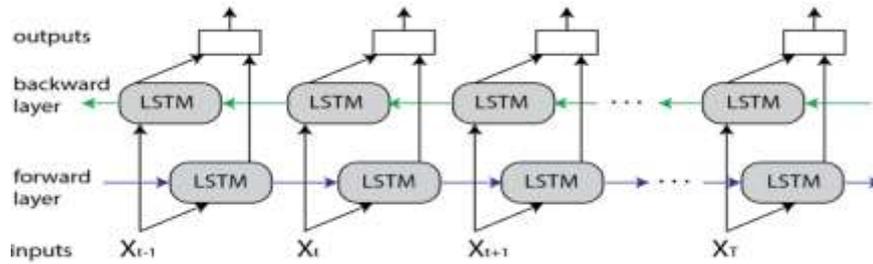
Fold 1	Data Tes	Data Validasi	Data Latih	Data Latih	Data Latih	Data Latih
Fold 2	Data Latih	Data Tes	Data Validasi	Data Latih	Data Latih	Data Latih
Fold 3	Data Latih	Data Latih	Data Tes	Data Validasi	Data Latih	Data Latih
Fold 4	Data Latih	Data Latih	Data Latih	Data Tes	Data Validasi	Data Latih
Fold 5	Data Latih	Data Latih	Data Latih	Data Latih	Data Tes	Data Validasi
Fold 6	Data Validasi	Data Latih	Data Latih	Data Latih	Data Latih	Data Tes

← (±) 16.7% →

Gambar 3 Proses Data K-Fold Cross Validation

### 2.2.4 Jaringan Syaraf Tiruan (JST)

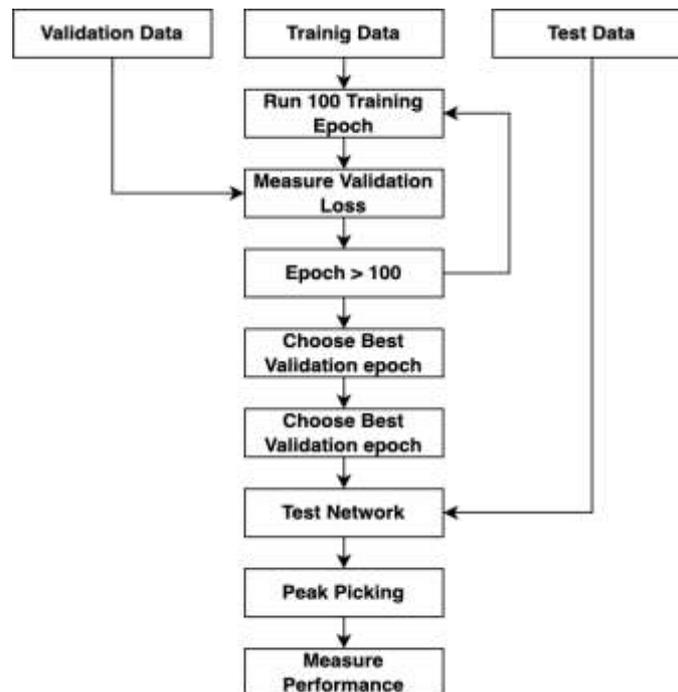
Jaringan Syaraf Tiruan (JST) berperan pada proses pelatihan data *onset*. *Multi Layer Perceptron* (MLP) adalah arsitektur yang paling umum digunakan. MLP termasuk ke dalam *feed forward network* (FFN) dimana hanya ada koneksi ke depan. Pengembangan model tersebut dengan menambahkan *cyclic connection* sehingga memiliki koneksi dengan memori pendek ke belakang. Model tersebut dinamakan *Recurrent Neural Network* (RNN). RNN digunakan dalam penelitian sebelumnya dengan menggunakan arsitektur *Elman Network* [5]. Jaringan RNN memiliki kelemahan dalam mempelajari *long term dependency* yang dinamakan *vanishing gradient* [9].



Gambar 4 Ilustrasi BLSTM

Dalam banyak penelitian, Model LSTM diklaim mampu mengatasi masalah *vanishing gradient* [9]. LSTM Terdiri dari dua jenis, yaitu Unidirectional dan Bidirectional. Unidirectional LSTM (LSTM) memproses masukkan dalam satu arah dengan cara mengalirkannya dari *past* (informasi lampau) ke *future* (informasi selanjutnya). Bidirectional LSTM (BLSTM) Memproses masukkan dalam dua arah dengan cara mengalirkannya *past* (informasi lampau) ke *future* (informasi selanjutnya) dan juga sebaliknya dari *future* ke *past*. BLSTM diklaim mampu memahami context lebih baik dari *Unidirectional LSTM* karena memiliki akses informasi dari *past* maupun *present*. Namun BLSTM memiliki kekurangan yaitu waktu pelatihan yang lebih lama disbanding dengan *Unidirectional LSTM*. Ilustrasi BLSTM ditunjukkan pada Gambar 4.

Penelitian ini menggunakan BLSTM 2 layer dengan masing-masing layer memiliki 100 unit neuron. Dikarenakan masalah deteksi *onset* adalah masalah klasifikasi *binary*, model ini menggunakan metode *Binary Cross Entropy* dalam mengukur *loss* pada model. Serta menggunakan *optimizer* berupa *stochastic gradient descent* dengan mengacu pada [8]. Proses pembelajaran seperti ditunjukkan pada Gambar 5 dimulai dari proses pelatihan dengan menggunakan data latih. Data akan dilatih sebanyak 100 langkah (*epoch*) dan setiap *epoch* dihitung nilai *loss* dan *validation loss* menggunakan data validasi. Setelah *epoch* terselesaikan, model dengan validasi terbaik akan dipilih untuk kemudian dilakukan evaluasi untuk mengukur kinerja model.



Gambar 5 Alur Kerja Model

### 2.2.5 Peak Picking

Dikarenakan model khusus untuk instrumen gamelan, maka proses seleksi kandidat *onset* menggunakan proses *thresholding* yang tetap. Proses ini akan menggunakan Fungsi baru  $G(t)$  yang dibangun menggunakan nilai ambang 0.25, yang mempertahankan nilai dari  $b(t)$  jika lebih besar dari  $\theta$  seperti ditunjukkan pada persamaan 3.

$$G(n) = \{b(t) \text{ if } b(t) > 0 \text{ 0 otherwise} \quad (3)$$

### 2.2.6 Pegujian (Evaluasi)

Fungsi evaluasi membutuhkan dua argumen wajib yaitu data prediksi dan data anotasi atau data *ground truth*. Kedua data tersebut berisi *onset* dalam satuan *milliseconds*. Proses evaluasi menggunakan modul pada pustaka *madmom*. Parameter yang perlu diperhatikan adalah parameter *combine* dimana fungsi hanya akan mendeteksi 1 saja onset jika masih dalam rentang waktu tersebut [9]. Parameter *combine* diisi dengan nilai 0.03 (nilai default) yang umum digunakan dalam proses deteksi onset [10]. Nilai 0.03 adalah nilai yang cukup aman karena hampir mustahil manusia secara normal bisa mendengar ketukan onset yang berbeda dalam rentang jarak tersebut. Sedangkan parameter *window* berfungsi sebagai nilai toleransi jarak onset dikatakan sama antara prediksi sama data *ground truth*. Nilai 0.025 menjadi nilai yang cukup aman dengan alasan yang sama dengan pemberian nilai *combine*.

## 3. HASIL DAN PEMBAHASAN

Hasil dari pengujian terdiri dari beberapa nilai yaitu *true positive* (TP) titik onset yang terdeteksi dengan benar, *false positive* (FP) titik bukan onset yang terdeteksi sebagai onset, *false negative* (FN) titik onset yang terlewatkan. Ketiga nilai tersebut menjadi perhitungan untuk mendapatkan nilai dari *Precision* (P), *Recall* (R), dan *F-measure* seperti ditunjukkan dalam persamaan 4, 5 dan 6.

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$F - \text{measure} = \frac{2PR}{P + R} \quad (6)$$

Hasil evaluasi berdasarkan keterlibatan instrumen ditunjukkan pada Tabel 2 Rata-rata instrumen yang paling pendek adalah instrument bonang yaitu 0.25 detik, sedangkan yang paling Panjang adalah instrument demung dengan rata-rata 0.71 detik.

Tabel 2 Hasil evaluasi berdasarkan Instrumen yang terlibat

Nama Instrument	Total Onset	Rata-Rata Interval	LR: 0,01			LR: 0,05		
			TP	FP	FN	TP	FP	FN
Saron	428	0,7180	426	3	2	426	7	2
Demung	187	0,8145	184	5	3	184	9	3
Peking	134	0,5128	134	1	1	134	1	1
Bonang	359	0,2528	321	12	38	327	20	32
Saron, Peking	84	0,4134	79	2	5	81	2	3
Bonang, Demung	114	0,3234	105	1	9	108	4	6
Bonang, Saron	212	0,3972	192	15	20	191	15	21
Bonang, Peking	133	0,2640	120	1	13	125	1	8
Total	1651		1561	40	91	1576	59	76

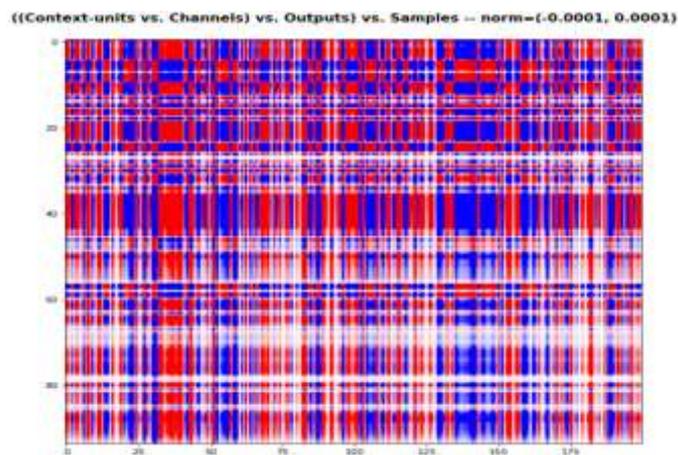
Dari tabel 2 disajikan ketiga nilai parameter evaluasi untuk masing-masing instrument dan *learning rate*. Tidak ada perbedaan yang signifikan pada nilai F-Measurenya dari kedua *learning rate* seperti ditunjukkan pada Tabel 3 Perbedaan ada pada nilai *precision* dan *recall* dimana *leaning rate* 0,01 sedikit unggul pada bagian *precision* sedangkan *leaning rate* 0,05 unggul pada bagian *recall*. Hal itu berkaitan dengan informasi pada Table 2 dimana *learning rate* 0,01 memiliki nilai FP yang lebih sedikit dan FN lebih banyak sedangkan *leaning rate* 0,05 memiliki nilai FN yang lebih sedikit dan FP lebih banyak.

Tabel 3 Hasil evaluasi secara umum

Learning rate	Tvpe	Precision	Recall	F-Measure
0.01	SUM	0.975	0.945	0.960
	MEAN	0.977	0.967	0.970
0.05	SUM	0.964	0.954	0.959
	MEAN	0.968	0.972	0.969

### 3.1 Pembelajaran Long Temporal Dependency

Berdasarkan Tabel 2, model yang telah dibuat bekerja dengan baik pada instrumen yang rata-rata interval onsetsnya relative jauh yaitu  $\geq 0,5$  detik yang ada pada Saron, Demung, dan Peking. Hal tersebut berkaitan dengan kemampuan model dalam mempelajari *Long-term Dependency* yaitu kemampuan jaringan dalam menangkap informasi historis jangka panjang. Gambar 6 menunjukkan model secara umum dapat mempertahankan nilai gradient pada saat proses pelatihan dimana hampir setiap *channels* dan *unit* memiliki nilai merah (sangat tinggi) biru (cukup tinggi) dan hanya sebagian kecil yang memiliki nilai putih (rendah). Gambar tersebut di plot menggunakan *norm* (-0.0001, 0.0001) dengan menyesuaikan jumlah *unit* yang terlibat pada jaringan BLSTM [11].



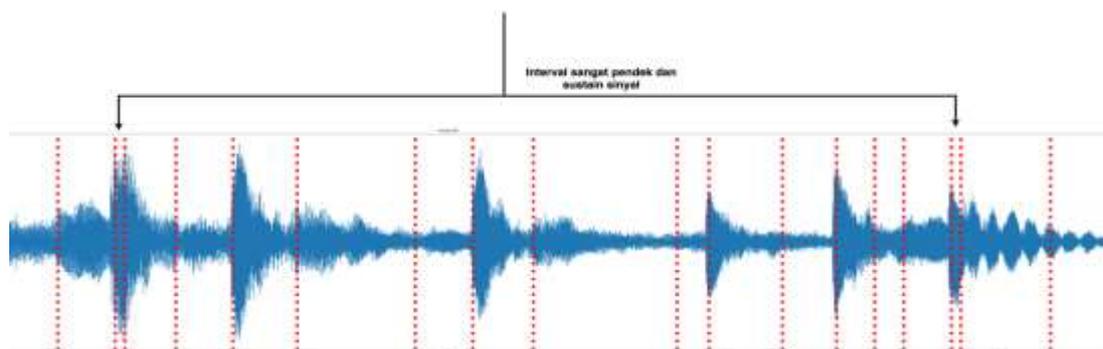
Gambar 6 Gradien pada Model

### 3.2 Pembelajaran Long Temporal Dependency

Sedangkan kesalahan pendeteksian onset sebagian besar terjadi saat ada keterlibatan instrumen boang di dalamnya dimana pada Tabel 2 ditunjukkan bahwa instrumen tersebut menyumbang proporsi *False Negatif* sebesar 87%, *False Positif* sebesar 73% pada *learning rate* 0,01 dan *False Negatif* sebesar %, *False Positif* sebesar 68% pada *learning rate* 0,05. *Magnitude* instrumen bonang lebih kecil dibanding instrumen lain seperti ditunjukkan pada Gambar 7. Serta cara memainkan instrumen bonang membuat daya sustain setelah ketukan lebih panjang dari instrumen lain, ditambah intervalnya yang pendek membuat onset dan non onset tidak memiliki fluktuasi *magnitude* yang signifikan seperti ditunjukkan pada Gambar 8.



Gambar 7 Amplitudo instrumen bonang



Gambar 8 Amplitudo instrumen bonang

#### 4. KESIMPULAN

Model dapat mengatasi mempelajari secara efektif *onset* yang intervalnya relative jauh. Hal itu ditunjukkan pada nilai kesalahan pendeteksian yang kecil pada instrument Saron, Demung, dan Peking. Hal tersebut menunjukkan kemampuan model dalam mempelajari *long-term dependency*.

Sebaliknya belum mampu mengatasi permasalahan instrumen dengan *magnitude* yang relatif kecil yaitu Bonang. Selain itu penelitian ini menemukan masalah lain yang belum teratasi yaitu daya *sustain* instrumen bonang yang lebih panjang dan interval yang sangat pendek membuat fluktuasi sinyal pada onset kurang tinggi sehingga banyak *onset* gagal dipelajari. Saran untuk penelitian selanjutnya untuk mengatasi masalah ini adalah menggunakan metode *adaptive thresholding*.

#### UCAPAN TERIMA KASIH

Terima kasih kepada rekan di SMK Negeri 1 Kasihan atas penyediaan data audio.

#### DAFTAR PUSTAKA

- [1] M. Mounir , P. Karsmakers and T. V. Waterschoot, "Musical note onset detection based on a spectral sparsity measure," EURASIP Journal on Audio, Speech, and Music Processing. 2021, Article no. 30, 2021 [Online]. Available: <https://asmp-urasipjournals.springeropen.com/articles/10.1186/s13636-021-00214-7> [Accessed 22-Nov-2022]

- [2]. B. Stasiakand, J. Mońko, and A. Niewiadomski, “NOTE ONSET DETECTION IN MUSICAL SIGNALS VIA NEURAL-NETWORK-BASED MULTI-ODF FUSION” *Int. J. Appl. Math. Comput. Sci.*, 2016, Vol. 26, No. 1, 203–213].
- [3]. E. Benetos, S. Dixon, Z. Duan, S. Ewert, Automatic music transcription: an overview. *IEEE Signal Process. Mag.* 36(1), 20–30 (2019).
- [4]. Risnandar., 2018, Pelarasan Gamelan Jawa.
- [5]. D. K. Sari, D. P. Wulandari. And Y. K. Suprpto, “Training Performance of Recurrent Neural Network using RTRL and BPTT for Gamelan Onset Detection”, International Conference on Electronics Representation and Algorithm (ICERA 2019)
- [6]. A. Rizal, R. Hidayat & H. A. Nugroho, “COMPARISON OF DISCRETE WAVELET TRANSFORM AND WAVELET PACKET DECOMPOSITION FOR THE LUNG SOUND CLASSIFICATION”, *Far East Journal of Electronics and Communications*, vol. 17, p.1065-1078, 2017
- [7]. B. Faghih, S. Chakraborty, A. Yaseen and J. Timoney, “A New Method for Detecting Onset and Offset for Singing in Real-Time and Offline Environments”, *Appl. Sci.*, vol. 12, p.7391, 2022 [Online]. Available: <https://www.mdpi.com/2076-3417/12/15/7391>. [Accessed: 22 Nov 2022]
- [8]. A. Schindler, T. Lidy and S. Böck, “Deep Learning for Music Information Retrieval”, 2018 [Online]. Available: [https://github.com/slychief/ismir2018\\_tutorial](https://github.com/slychief/ismir2018_tutorial), [Accessed 3-Aug-2022]
- [9]. J. Zhang, Y. Zeng, B. Starly. “Recurrent neural networks with long term temporal dependencies in machine tool wear diagnosis and prognosis”, *SN Appl Sci*, vol.3, p.442. 2021[Online]. Available: <https://doi.org/10.1007/s42452-021-04427-5>. [Accessed: 22 Nov 2022]
- [10]. Schindler, A., Lidy, T., & Böck, S., 2018, Deep Learning for Music Information Retrieval, [https://github.com/slychief/ismir2018\\_tutorial](https://github.com/slychief/ismir2018_tutorial), diakses pada 3 Agustus 2022
- [11]. J. Muradeli, “See-RNN”, 2019 [Online], Available: <https://github.com/OverLordGoldDragon/see-rnn>, [Accessed: 22 Nov 2022]