ScienceDirect

Procedia Computer Science 00 (2023) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Computer Science and Computational Intelligence 2023

DeepLyric: Predicting Music Emotions through LSTM-GRU Hybrid Models with Regularization Techniques

Felicia Angelica[a,*], Romington Hydro[a], Zefanya Delvin Sulistiya[a], Yohan Muliono[a], Simeon Yuda Prasetyo[a]

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

Abstract

Music is a powerful medium that has the potential to evoke strong emotional responses in humans. Researchers have investigated the relationship between music and emotions and the growing interest in developing precise techniques for recognizing emotions in music. Growing interest among researchers in developing more precise techniques for recognizing emotions in music has evoked numerous studies to be conducted. The researchers trained LSTM and GRU (Gated Recurrent Unit) hybrid model on a dataset of 1,160 songs tagged with sadness, compassion, and tension. They were able to attain a test accuracy of 72.51%, although overfitting was discovered because of insufficient training data, a choice of dropout, and learning rate. The research proves the potential of machine learning methods for musical emotion recognition and recommends the use of regularization techniques to alleviate overfitting. It may be possible to create software that can recognize and respond to emotional states in musical contexts because of the research, which contributes to the growing body of knowledge on the connection between music and machine learning.

© 2023 The Authors. Published by ELSEVIER B.V. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 8th International Conference on Computer Science and Computational Intelligence 2023

Keywords: Song Lyrics; Music Emotion Recognition; GRU; LSTM; Natural Language Processing; Hybrid Model

1. Introduction

Music is an effective medium with plenty of positive impacts on the body and mind. The direct connection between music and intense emotional experiences is highlighted by the fact that research has shown that the portion of the human brain that processes music is tightly linked to the area where emotions are expressed (1). Strong emotions can be evoked by music in a variety of ways, including the expression of internal sensations, the sensation of goosebumps, the capacity to make listeners cry, or the ability to experience an emotional state alongside the composer or performer

1877-0509 © 2022 The Authors. Published by ELSEVIER B.V. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 7th International Conference on Computer Science and Computational Intelligence 2022

(2). Numerous studies have been conducted about music's emotional power, including ones that look at how it can both express and evoke emotions as well as how it can be judged aesthetically (3). Therefore, there is growing interest among researchers in developing more precise techniques for recognizing emotions in music.

In recent years, the study of automatic music emotion recognition has expanded. Personalized playlists, music recommendation programs, music therapy interventions, and the use of music in commercials to induce feelings in viewers are just a few uses for being able to understand the emotional nature of music. In this context, the current study focuses on the use of machine learning methods to identify musical emotions within lyrics.

The study employs a hybrid model to categorize lyrics into several emotional states, such as grief, tenderness, and tension. This model combines long short-term memory (LSTM) and gated recurrent unit (GRU) cells. Prior studies have amply demonstrated the usefulness of Recurrent Neural Network (RNN) models in identifying both short-term and long-term dependencies in sequential data, such as song lyrics.

Lyrics are a sequential data, where the order of each words and lines plays crucial role for bring up meaning and emotion for a song. This is the cause why this study proposed to use LSTM and GRU model to recognize the emotion from a lyrics. LSTM and GRU are both the types of Recurrent Neural Networks (RNNs) that are designed to capture a sequential information or data.

The LSTM and GRU model have a gates as an internal mechanism. These gates can control what information to keep and what information needs to be thrown out by the model. By combining a different recurring layer from LSTM and GRU can enhance the model’s generalization and help to reduce the overfitting problem.

The proposed approach dissects the linguistic properties of song lyrics to appropriately classify them into one of three different emotional states. Songs that transmit warmth, sympathy, or intimacy are categorized as tender, whereas sad songs reflect feelings of sorrow, grief, or melancholy. Tension in lyrics is defined as feelings of suspense, anxiety, or anticipation. The approach provides insightful data on the emotional resonance of music and how it affects listeners.

The goal of this study is to develop a novel and reliable model for recognizing emotions in music based on lyrics. The results of this research may expand our knowledge of the emotional content of song lyrics and how they relate to musical form. Potential applications for the suggested technique of employing LSTM-GRU hybrid models for music emotion identification based on lyrics are in the psychology and music fields.

2. Related Work

The ability of music to express and communicate emotions has long attracted scholars, particularly given the growing importance of emotional intelligence. Understanding emotions in music has substantial effects on emotional health and has useful uses in marketing, therapy, and music production. This review of the literature analyses recent studies that have been published during the last five years and highlights significant findings, their relevance to the subject, and directions for future study. These studies looked at how music affects emotions with a particular emphasis on song lyrics. The findings may provide insights into how music expresses emotions and propose future study.

The effects of lyric features on the perceived arousal and valence of music were explored, and music emotion identification models utilizing audio and lyric data were proposed in the article "Using machine learning analysis to interpret the relationship between music emotion and lyric features" by Xu et al. According to the study, audio qualities are more significant than lyric aspects when it comes to the impression of musical emotion (4).

In a study by Jia (2022), a CNN-LSTM model was proposed for research in emotion recognition based on music lyrics. The authors used a hybrid model to process the model and gain accuracy. The authors used WORD2VEC and TFIDF for the feature extraction. Then after the feature extraction the model will be processed by CNN-LSTM model and gain an accuracy of 84.8%. Another approach used a Bi-GRU model with an attention layer named a SEER model to process the text and gain accuracy 83.67% (5).

In another study, lyrics were divided into 3 groups according to 3 different emotions: sadness, tension, and tenderness. The researchers found that their model performed more effectively in correctly categorizing song lyrics than SVM (Support Vector Machine) and decision tree models. Another study looked at how to divide up the emotions expressed in song lyrics into two groups: valence and arousal. The authors used 2 models and achieved high accuracy in classifying lyrics (6).

Another study by Minho (2022), an SVM model was proposed for the research in emotion recognition based on music lyrics using feature selection by Partial Syntactic Analysis. Firstly, the dataset is being pre-processed then the model will be processed by the SVM model and gain an accuracy 60.8% for 2829 lyrics songs and 58.8% for 425 lyrics songs (7).

Another study about the emotion recognition based on text is proposed with a hybrid model using Bi-GRU, GRU, and CNN model to process the dataset model. The authors used TFIDF and WORD2VEC for the feature extraction to pre-process the dataset before it be processed by the hybrid model. After the feature extraction using TFIDF and WORD2VEC, the model is processed by the hybrid model using Bi-GRU, GRU, and CNN and gains an accuracy 80.11% (8).

A research that wants to try to determine emotion of a song and try to understand how deep learning affect classification process uses Long Short-Term Memory (LSTM), RNN, K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and CNN. It uses MoodyLyric database for song index, artist, title, and emotion set and also Genius and SongLyrics for the lyrics. The research shows that the optimal learning rate is at 0.00006 and the most effective model is Bi-LSTM with 91,08% accuracy (9).

Overall, these studies demonstrate the effectiveness of hybrid models for music emotion recognition based on lyrics. The combination of LSTM and GRU cells in an RNN architecture may allow for the effective capture of short-term and long-term dependencies in sequential data, which is essential for accurately classifying the emotional content of song lyrics.

3. Methodology

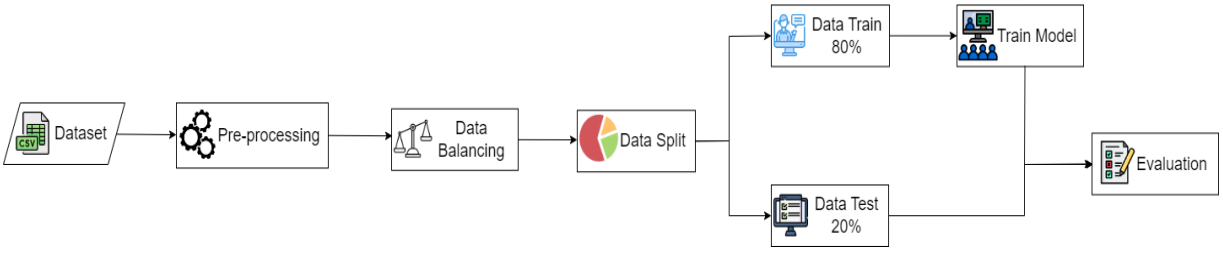

Fig. 1. Research Workflow

The workflow in Fig.1 represents the steps to do the research for our experiment. The first step is to collect the data for our experiment. After finding the dataset, we preprocess the data before we pass it to the model to gain accuracy.

3.1. Dataset

In this study, we conducted an experiment on a dataset of 1,160 songs, each of which was assigned one of three emotions: sadness, tenderness, or tension. The dataset occurred from the repository of Diptanu (6). To guarantee the validity of the results we obtained, we separated the dataset into training, validation, and testing sets in 80%, 10%, and 10% subsets, respectively. The LSTM-GRU hybrid model was employed in our research. These models were trained on the training set and assessed using sparse categorical entropy and the Adam optimizer on the validation set. The objective was to select the model that performed the best and had the highest accuracy on the validation set for the testing set's final evaluation. The dataset label distribution is shown in Fig 2.

We use LSTM and GRU because these models have gates as an internal mechanism. These gates can control what information to keep and what information to be thrown out. Research by Minho (7), the author used SVM to classify the emotion of songs. Another research by Jia (10), the author used hybrid model using LSTM and CNN (Convolutional Neural Network) model to classify the emotion of songs. We compare the models, and the best model is the hybrid model with an accuracy 82.6% compared to the SVM model which gain an accuracy 60.8% for text classification.

We use sparse categorical entropy for the losses because this loss is for a sample that has only a single label. For our case, we only use a single label not a multilabel. We use Adam optimizer because this optimizer has a faster computation time and minimizes the cost of the loss function rather than the other optimizer.

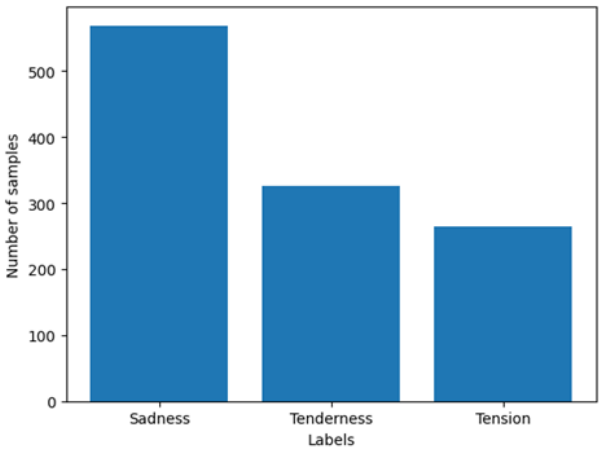

Fig. 2. Dataset Distribution for Each Songs Label

3.2. Preprocessing

Stop words, punctuation, and other special characters were deleted from the data during the preprocessing stage, and the remaining words were then translated into numerical representations using previously trained word embeddings. The minority classes in the dataset were oversampled using the RandomOverSampler approach to ensure that the dataset was balanced across the three emotion categories. RandomOverSampler only works in data training process. RandomOverSampler operates by randomly duplicating the data points of minority class to be balanced by the major class (11). Padding was applied to ensure that all sequences were of the same length, and early stopping was used to avoid overfitting and ensure that the model generalizes well to new data. In this case we use pre-padding where empty tokens are added into the beginning of the sentence to make all sentences have the same length (12).

3.3. Data Balancing

To ensure that the dataset was balanced across the three emotion categories, the minority classes in the dataset were oversampled using the RandomOverSampler method (13). Early stopping was employed to prevent overfitting, ensuring that the model generalizes well to new data, and apply padding to guarantee that all sequences had the same length.

3.4. Hybrid Model

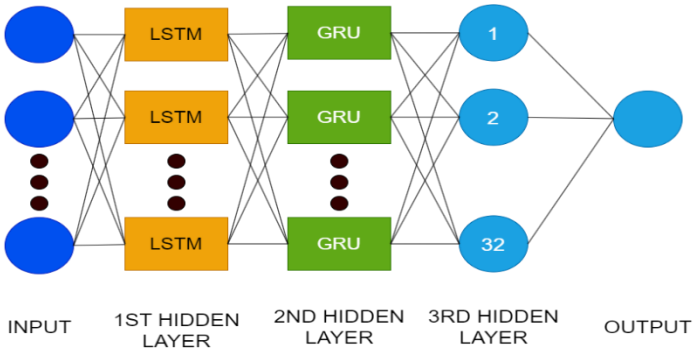

Fig. 3. LSTM-GRU Hybrid Model Structure

The authors used dropout regularization to avoid overfitting and created a hybrid LSTM-GRU model with one layer of LSTM cell and one layer of GRU cells. The categorical cross-entropy loss function was used to train the model, and the SoftMax activation function was utilized for predicting the probable outcome of each emotion type. The Adam optimization method was applied to improve the model's accuracy (14).

3.5. Evaluation Metrics

The effectiveness of the suggested model was assessed using parameters like accuracy, F1 score, precision, and recall (15,16). Its results were contrasted with those of other innovative algorithms for deciphering musical emotions from lyrics. The results indicate that the suggested hybrid model with LSTM-GRU cells performs well in identifying musical emotions from lyrics. The proposed methodology can be used in many different contexts, such as music recommendation systems, personalized playlists, and music therapy interventions.

4. Result and Discussions

In this test we use Google Colab to run our model, we use 4 parameters, LSTM, GRU, the hybrid model with dropout rate of 0,2 and another hybrid model with the dropout rate of 0,5. Here are the results of our experiment:

Model	Metrics			
	Accuracy	F1-score	Precision	Recall
LSTM	0.61206	0.31941	0.40720	0.36394
GRU	0.61206	0.25311	0.20402	0.33333
Hybrid [LSTM+GRU] (Dropout = 0.2)	0.71929	0.71431	0.71649	0.71434
Hybrid [LSTM+GRU] (Dropout = 0.5)	0.72514	0.72325	0.72445	0.72280

Table 1. Experimental Results

We use dropout as a regularization to prevent an overfitting problem. The dropout can prevent overfitting problem by randomly dropping out some model unit in the neural network during the training process. Dropout is an effective regularization in deep learning.

From Table 1, we can see that for the hybrid model with dropout 0.5 got 72.51% on accuracy, 72.33% f1-score, 72.45% precision, and 72.28% recall. As we can see, the hybrid model with 0.5 dropout gains a higher accuracy than the 0.2 dropout. It is because higher dropout will result in more regularization to prevent an overfitting problem. An optimal dropout depends on the complexity of the model and the size of the datasets.

Based on Table 1, the LSTM and GRU model gain less accuracy than the hybrid model. It is because hybrid models use 2 generalizations. The LSTM and GRU have their own regularization. Generalization is used to reduce the risk of overfitting. So, the hybrid model gains more accuracy than the single model because there are 2 generalizations to reduce the risk of overfitting.

From Fig. 4, we can see that 37 data labelled with Sadness is correct, 43 data labelled with Tension is correct, and 41 data labelled with Tenderness is correct. While there are 17 data supposed to be labelled with Sadness but labelled with Tension and 1 data supposed to be labelled with Tenderness but labelled with Tension. The darker the space the less accuracy of the output.

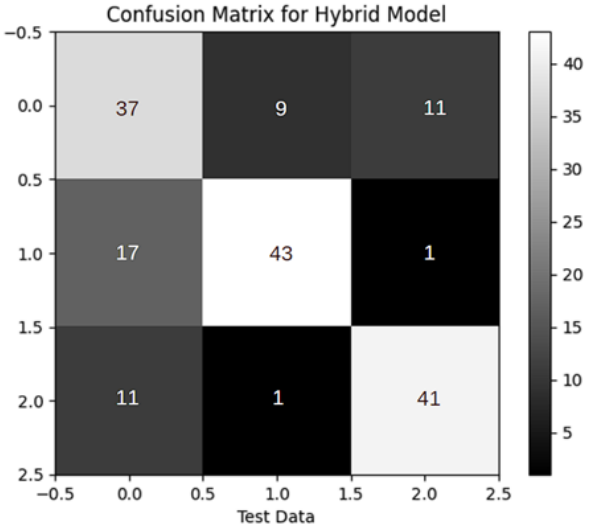

Fig. 4. Confusion Matrix Hybrid (Dropout-0.5)

Confusion matrix is used to evaluate the performance of the model. The row represents the actual classes, the column represents the predicted classes, The diagonal represents the number of the correct predictions while the off-diagonal element represents the number of incorrect predictions.

The outcomes of our study show that the best performing model obtained an accuracy of 72.51% on the testing set. However, we found that the model was overfitting during the experiment, as shown by the graph used to assess the model's accuracy. Despite this, our findings show the utility of LSTM-GRU hybrid model for music emotion identification tasks.

Overall, by offering information about the effectiveness of LSTM and GRU hybrid models on a large dataset of songs labelled with emotions, our study contributes to music emotion recognition. Our findings emphasize the significance of comprehensive model evaluation and model selection to guarantee the validity and generalizability of outcomes. The performance of these models might be enhanced, and the overfitting problem could be addressed, through additional study.

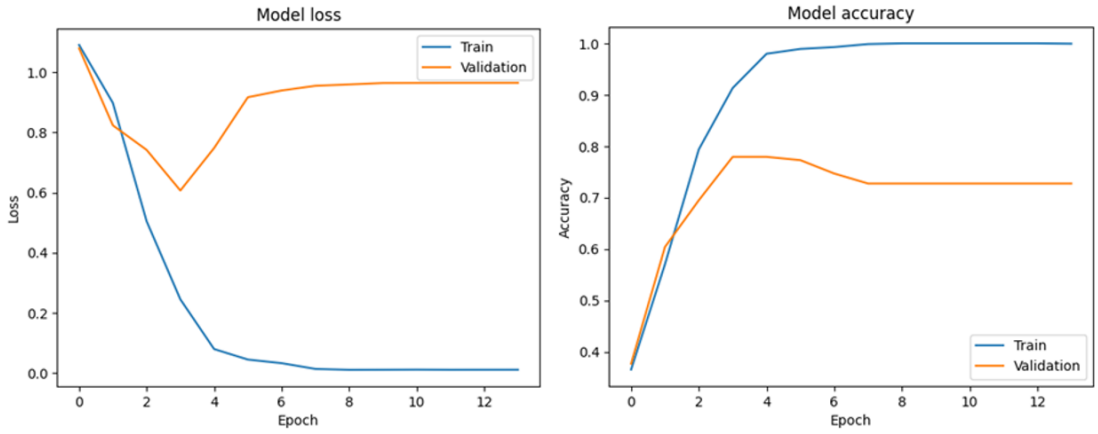

Fig. 5. Model Loss Graphics (Left) and Model Accuracy Graphics (Right)

The graphic shows that machine learning is overfitting. Overfitting occurs when the validation set number is bigger than the training set. Many factors caused machine learning to be overfitting. The first factor is due to a lack of adequate training data. The second factor is the chosen dropout and learning rate which can also cause the overfit

problem. The third factor is about the model complexity, a complex model can cause an overfitting problem to a simpler model. The fourth factor is the data pre-processing which can cause an overfitting problem.

There are many ways to overcome the overfitting graph. The first thing to do is increase the training dataset so when a complex model is implemented to the training data, the graph will not be overfitting. Secondly, use proper regularization. A proper regularization such as L1 or L2 regularization could be utilized before feeding the data into the models to overcome the overfitting graph. Thirdly, use an early stopping to monitor the validation loss during the training set and stop the training process when the validation loss stops improving. Lastly, is using a dropout to temporarily remove random neurons from the model during training process (17).

Our models have managed to outperform other work with SVM model (7), where the SVM model only gain 60,8% accuracy while our model gain 72.51% accuracy. But our model still cannot outperform other hybrid models with a CNN-LSTM model (10) that gains 84.8% accuracy while our model gains 72.51% accuracy. This is caused by the dataset that we used and the pre-processed before the modelling occurred. Our dataset only consists of 1.160 lyrics while other researchers with a CNN-LSTM model have 5.286 data.

5. Conclusion

Music emotion recognition is an interesting and fast expanding subject of research in machine learning, as music is a vital aspect of human emotions in everyday life. For this study, the researchers used a dataset of 1,160 songs to train our models on three emotion labels: sadness, tenderness, and tension. The dataset was divided into three sets for the machine learning process: training, validation, and testing. We examined the dataset using LSTM-GRU hybrid model for training and attained a test accuracy of 72.51% with a dropout of 0.5 and a learning rate of 0.001. However, we identified overfitting in our training data, which we primarily attribute to a lack of adequate training data. The chosen dropout and learning rate may also have contributed to this issue. Regularization techniques like L1 or L2 regularization could be utilized before feeding the data into the models to overcome this problem.

Our findings show that machine learning approaches can be used to recognize musical emotions, although more study is required to overcome overfitting problems and boost accuracy. Overall, our research adds to the expanding body of knowledge on the relationship between music and machine learning and may have repercussions for the creation of software that can effectively detect and react to emotional states in musical contexts. In the future we intend to use our proposed method for other datasets to check the consistency of the obtained result from our research.

References

1. Subramanian RR, Ram KA, Sai DL, Reddy KV, Chowdary KA, Reddy KDD. Deep Learning Aided Emotion Recognition from Music. In: 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS). IEEE; 2022. p. 712–6.
2. Gomez-Canon JS, Cano E, Eerola T, Herrera P, Hu X, Yang YH, et al. Music Emotion Recognition: Toward new, robust standards in personalized and context-sensitive applications. IEEE Signal Process Mag. 2021 Nov;38(6):106–14.
3. Juslin PN. Musical Emotions Explained. Oxford University Press; 2019.
4. Xu L, Sun Z, Wen X, Huang Z, Chao C ju, Xu L. Using machine learning analysis to interpret the relationship between music emotion and lyric features. PeerJ Comput Sci. 2021 Nov 15;7:e785.
5. Jia X. Music Emotion Classification Method Based on Deep Learning and Improved Attention Mechanism. Comput Intell Neurosci. 2022 Jun 20;2022:1–8.
6. Sarkar D. Detecting emotions in lyrics. 2020 [cited 2023 May 8]; Available from: https://www.diptanu.com/data/lyrics_emotion_detection.pdf

7. Kim M, Kwon HC. Lyrics-Based Emotion Classification Using Feature Selection by Partial Syntactic Analysis. In: 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence. IEEE; 2011. p. 960–4.
8. Bharti SK, Varadhaganapathy S, Gupta RK, Shukla PK, Bouye M, Hingaa SK, et al. Text-Based Emotion Recognition Using Deep Learning Approach. Comput Intell Neurosci. 2022 Aug 23;2022:1–8.
9. Jiddy Abdillah, Ibnu Asror, Yanuar Firdaus Arie Wibowo. Emotion Classification of Song Lyrics using Bidirectional LSTM Method with GloVe Word Representation Weighting. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi). 2020 Aug 20;4(4):723–9.
10. Jia X. Music Emotion Classification Method Based on Deep Learning and Improved Attention Mechanism. Comput Intell Neurosci. 2022 Jun 20;2022:1–8.
11. Hayaty M, Muthmainah S, Ghufran SM. Random and Synthetic Over-Sampling Approach to Resolve Data Imbalance in Classification. International Journal of Artificial Intelligence Research. 2021 Jan 5;4(2):86.
12. Dwarampudi M, Reddy NVS. Effects of padding on LSTMs and CNNs. 2019 Mar 18;
13. Wongvorachan T, He S, Bulut O. A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. Information. 2023 Jan 16;14(1):54.
14. Ullah I, Raza B, Ali S, Abbasi IA, Baseer S, Irshad A. Software Defined Network Enabled Fog-to-Things Hybrid Deep Learning Driven Cyber Threat Detection System. Security and Communication Networks. 2021 Dec 3;2021:1–15.
15. Blagec K, Dorffner G, Moradi M, Samwald M. A critical analysis of metrics used for measuring progress in artificial intelligence. 2020 Aug 6;
16. Kaya, Bilge. Deep Metric Learning: A Survey. Symmetry (Basel). 2019 Aug 21;11(9):1066.
17. Ying X. An Overview of Overfitting and its Solutions. J Phys Conf Ser. 2019 Feb;1168:022022.