

## Implementasi SuperTML Untuk Klasifikasi Genre Musik Indonesia

**Joni Bastian**

Universitas Pembangunan Nasional “Veteran” Jawa Timur

**Made Hanindia Prami Swari**

Universitas Pembangunan Nasional “Veteran” Jawa Timur

**Andreas Nugroho Sihananto**

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

Korespondensi penulis: [19081010071@student.upnjatim.ac.id](mailto:19081010071@student.upnjatim.ac.id)

**Abstract.** *Music genres are becoming increasingly diverse, and many people listen to music because it has benefits such as refreshing, motivating or therapeutic. However, with the increasing number of genres, some listeners have a tendency towards the type of genre they like. In Indonesia itself, there are several popular music genres such as pop, folk, rock, indie and dangdut. Classification of music genres is an interesting topic when looking at this behaviour. Several approaches to classify popular music genres include audio and tabular data approaches. In this research, classifying music genres using an image approach by implementing SuperTML to change the form of tabular data into image form, which is then trained using a pre-trained CNN Densenet, succeeded in achieving an accuracy of 67%.*

**Keywords:** *SuperTML, Densenet, Genre, Musik*

**Abstrak.** Genre musik semakin beragam dan musik banyak didengarkan karena memiliki manfaat seperti refreshing, motivasi, ataupun terapi. Namun semakin banyaknya genre beberapa pendengar memiliki kecenderungan terhadap jenis genre yang disukai. Di Indonesia sendiri ada beberapa genre musik populer seperti pop, folk, rock, indie, dan dangdut. Klasifikasi genre musik menjadi topik yang menarik melihat perilaku tersebut. Beberapa pendekatan untuk mengklasifikasikan genre musik yang umum melalui pendekatan data audio dan tabular. Pada penelitian ini klasifikasi genre musik melalui pendekatan gambar dengan mengimplementasikan SuperTML untuk mengubah bentuk data tabular menjadi bentuk citra yang kemudian dilatih menggunakan *pre-trained* CNN Densenet berhasil mendapatkan akurasi dengan nilai 67%.

**Kata kunci:** *SuperTML, Densenet, Genre, Musik*

### LATAR BELAKANG

Dewasa ini, genre musik semakin beragam dan musik mulai banyak didengarkan sambil melakukan aktivitas lain karena musik juga dapat memberikan beberapa manfaat seperti refreshing, motivasi, dan terapi (Lidinillah Alfath dkk, 2022). Namun, dengan semakin beragamnya genre musik dan beberapa manfaat musik bagi pendengarnya, beberapa pendengar akan memiliki kecenderungan terhadap selera musiknya yang dapat mempengaruhi pendengar akan lebih suka pada genre tertentu. Beberapa hal yang dapat menjadi tolak ukur musik tertentu adalah musik yang memiliki genre Pop, Rock ataupun dangdut adalah timbre, rhythm, melody, harmony, dan pitch (Ayu & Giri, 2017).

Spotify merupakan aplikasi musik player yang memungkinkan pendengarnya menikmati musik atau podcast yang pendengar inginkan, kapan saja dan di mana saja. pendengar dapat mengakses jutaan lagu dan acara di perangkat apa pun. Perangkat juga dapat memilih lagu atau podcast pendengar sendiri, atau biarkan Spotify memilihkan sesuatu untuk pendengar. Pendengar juga dapat menjelajahi daftar putar teman, artis, dan selebriti favorit pendengar, atau mendengarkan stasiun radio berdasarkan suasana hati pendengar.

Spotify untuk pengembang membangun pengalaman bagi jutaan pecinta musik dengan pemutaran, personalisasi, dan banyak fitur lainnya. Spotify untuk pengembang memberikan akses bagi pengembang untuk menjelajahi fitur audio dan analisis trek yang mendalam. Pengembang dapat mencari tahu *danceability*, *instrumentalness*, dan *valence* dari trek favorit pendengar. Pengembang bisa mendapatkan analisis segmen , bar, nada dan banyak lagi lagu dengan platform spotify untuk pengembang. Pengembang dapat mengembangkan aplikasi dengan metadata musik tingkat lanjut.

Penelitian sebelumnya yang berjudul “Klasifikasi Genre Musik dengan Pendekatan *Convolutional Neural Network*” (Smith dkk., 2021), menggunakan pemrosesan ekstraksi audio untuk mendapatkan informasi spasial yang digunakan untuk mengklasifikasikan genre musik. Pada penelitian tersebut menunjukkan bahwa algoritma *Neural Network* dapat memberikan hasil yang baik pada kasus klasifikasi khususnya Genre Musik. Penelitian lain yang berjudul “SuperTML: Penyematan Kata Dua Dimensi untuk Prekognisi pada Data Tabular Terstruktur” yang diteliti oleh Sun Baohua dkk pada tahun 2019 memberikan gambaran bahwa kasus-kasus seperti *Deep Learning* atau *Neural Network* seringkali menggunakan penyematan satu dimensi ini akan cocok untuk kasus seperti citra, audio dan lainnya. Namun dengan SuperTML yang terinspirasi dari karakter super dan penyematan dua dimensi ini menjadikan masalah pada tabular data dapat diterapkan dengan algoritma-algoritma *Deep Learning* seperti *Neural Network*. Pada permasalahan penelitian terdahulu, penelitian kali ini akan menggunakan dataset yang diambil dari Spotify dengan API yang memberikan akses audio analisis tingkat lanjut. Pada penelitian ini Peneliti akan mengklasifikasikan genre musik secara umum di Indonesia yang ada pada platform Spotify.

## KAJIAN TEORITIS

### SuperTML

Metode SuperTML didasarkan pada analogi antara masalah *Tabular Meta Learning* (TML) dan tugas klasifikasi teks. Untuk setiap sampel yang diberikan dalam bentuk tabel, jika fitur-fiturnya diperlakukan seperti token data yang terstringifikasi, maka setiap sampel dapat

direpresentasikan sebagai penggabungan fitur-fitur yang telah diberikan tokenisasi. SuperTML terdiri dari dua langkah, langkah pertama adalah embedding dua dimensi. Langkah ini memproyeksikan fitur-fitur pada data tabular ke dalam gambar-gambar yang dihasilkan. Langkah kedua adalah menggunakan model-model CNN yang telah dilatih sebelumnya untuk fine-tuning pada gambar-gambar SuperTML yang dihasilkan.



**Gambar 1 Ilustrasi SuperTML (Sun Baohua dkk, 2019)**

SuperTML memiliki dua jenis yaitu SuperTML VF yang merupakan algoritma SuperTML untuk memberikan beberapa ruang lebih besar kepada fitur-fitur yang sangat penting, dan SuperTML EF yang memberikan ruang sama bagi setiap fitur sehingga bisa secara langsung disematkan pada SuperTML tanpa perlu kalkulasi fitur penting.

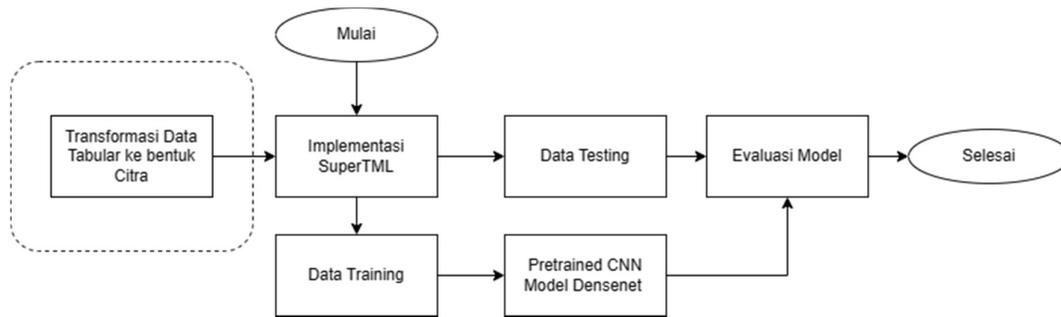
### DenseNet

Densenet atau *densely connected-convolutional networks* sangat mirip dengan ResNet namun memiliki perbedaan mendasar ResNet menggunakan metode aditif yang berarti mereka mengambil output sebelumnya sebagai input untuk layer masa depan, namun di DenseNet mengambil semua output sebelumnya sebagai input untuk layer masa depan (Huang Gao dkk, 2018).

### Neural Network

*Neural Network* atau Jaringan Saraf Tiruan dalam Bahasa Indonesia terinspirasi dari cara kerja jaringan saraf manusia. Seperti otak manusia, JST mampu memproses informasi dan belajar dari suatu objek melalui proses learning. JST mampu menyimpan pengetahuan yang diperoleh dan menggunakannya untuk memprediksi atau mengambil keputusan pada data baru. Penggunaan NN sangat luas, salah satunya dalam memprediksi kejadian di masa depan dan menyelesaikan masalah kompleks seperti otak manusia. NN mampu mempelajari pola-pola kompleks dalam data dan menghasilkan output yang dapat digunakan untuk membuat keputusan atau prediksi di masa depan (Sinaga, 2020).

## METODE PENELITIAN



**Gambar 2 Tahapan Penelitian**

### Implementasi SuperTML

Proses awal adalah mengubah bentuk data tabular menjadi bentuk citra. Pada proses ini data yang akan diubah bentuknya menjadi citra berjumlah 1160 baris data musik dengan jenis genre musik Indonesia berdasarkan spotify yang peneliti pilih yaitu:

**Tabel 1 Genre Musik**

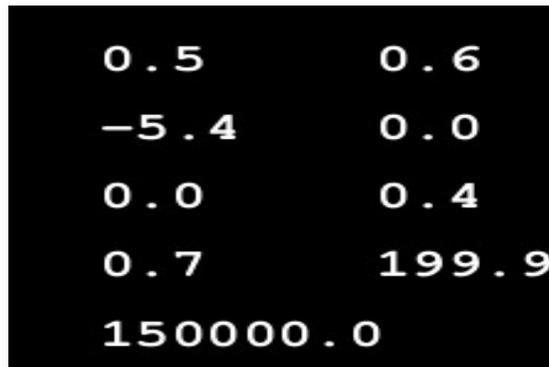
No	Genre
1.	Indonesian Pop
2.	Indonesian Folk
3.	Classic Indonesian Rock
4.	Dangdut
5.	Indonesian Indie

Berdasarkan genre yang Peneliti ambil yang sudah dilabeli berdasarkan Spotify kemudian Peneliti mengambil audio analisis dari setiap *track* musik dari 1160 musik tersebut. Beberapa fitur yang peneliti gunakan yaitu *danceability*, *energy*, *loudness*, *speechiness*, *instrumentalness*, *liveness*, *valence*, *tempo*, dan *duration\_ms* seperti terlihat pada tabel di bawah ini:

**Tabel 2 Fitur dan Genre Musik**

danceability	energy	loudness	speechiness	instrumentalness	liveness	valence	tempo	duration_ms	genre
0.508	0.956	-5.930	0.0835	0.00000	0.0315	0.636	149.710	240512	dangdut
0.481	0.877	-7.461	0.0467	0.00000	0.1080	0.618	171.975	212138	dangdut
0.618	0.776	-6.410	0.0282	0.00000	0.1270	0.842	133.986	300513	classic indonesian rock

Berikut merupakan *track*/musik yang sudah diimplementasikan metode SuperTML sehingga menjadi citra

**Gambar 3 Citra hasil implementasi SuperTML**

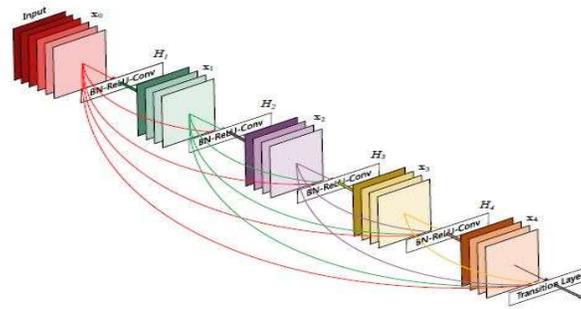
### Data Training

Pada data training dari total data keseluruhan berjumlah 1600 data citra 80% atau berjumlah 1280 citra dari total data dialokasikan ke dalam data training untuk disiapkan sebagai input model.

### Data Testing

Pada data testing dari total data keseluruhan berjumlah 1600 data citra 20% sisanya yang berjumlah 320 disiapkan untuk dialokasikan ke dalam data testing untuk disiapkan di tahap evaluasi model setelah model selesai di latih.

**Pretrained CNN Model DenseNet**



**Gambar 4 Ilustrasi Densenet**

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

**Gambar 5 Arsitektur Densenet**

Pada proses ini peneliti menggunakan arsitektur pretrained DenseNet121 dengan menggunakan weight imagenet kemudian peneliti tambahkan dengan 1 layer *neural network*. Peneliti juga menggunakan *loss function* sparse categorical crossentropy dan learning rate adamax.

**Evaluasi Model**

Pada proses ini data testing dilakukan pengujian terhadap model yang kemudian dievaluasi mengenai hasil prediksi dengan label sebenarnya dari data musik. Pada proses ini hasil dari evaluasi akan mempresentasikan performa model.

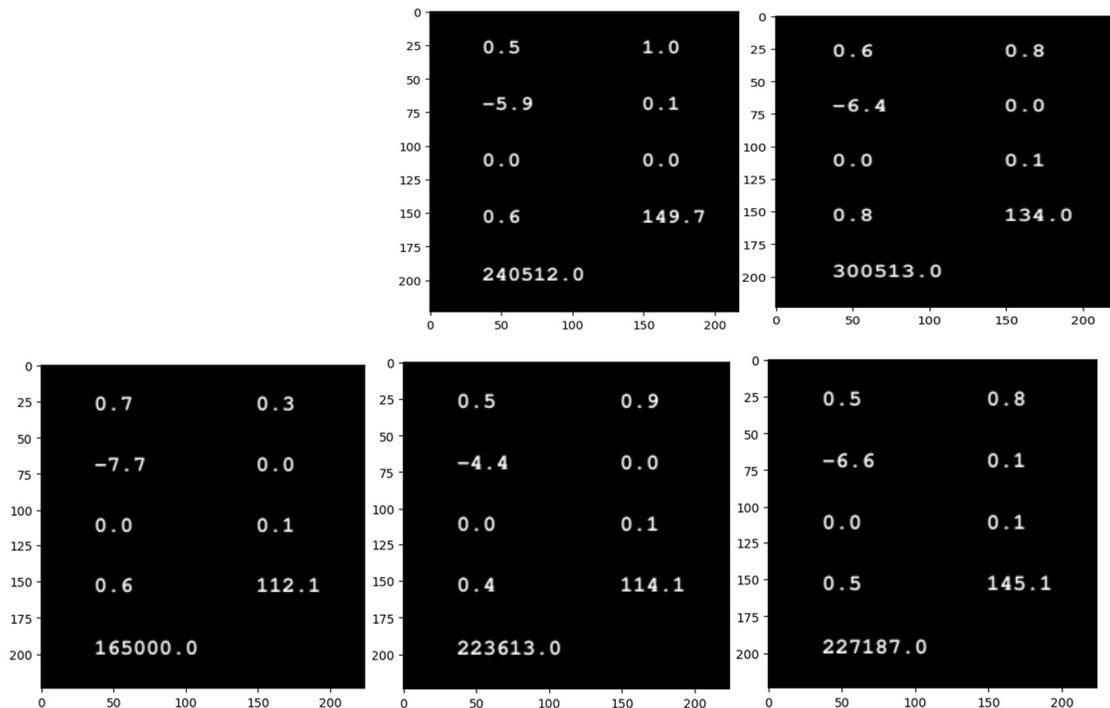
**HASIL DAN PEMBAHASAN**

Dataset yang telah dikumpulkan dari spotify berjumlah 1160 data yang berbentuk tabular kemudian dipisahkan antara fitur dan genre-nya. Berikut tabel musik dengan genre-nya yang berhasil dikumpulkan sejumlah 1160 data.

**Tabel 3 Fitur Audio Analisis dan Genrenya**

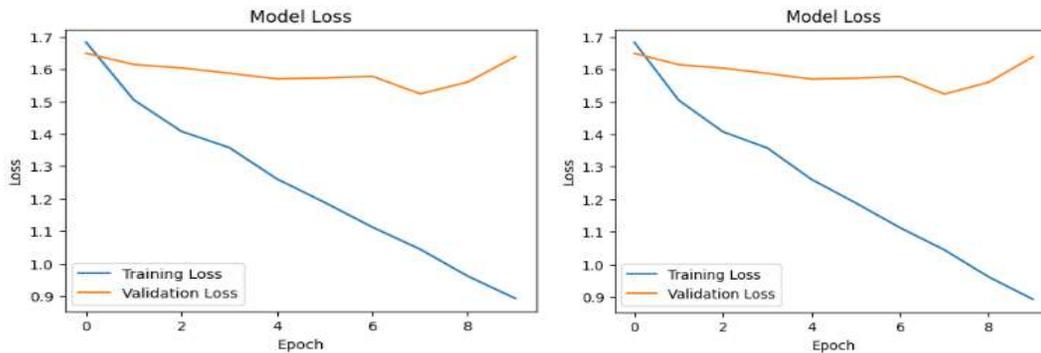
danceability	energy	loudness	speechiness	instrumentalness	liveness	valence	tempo	duration_ms	genre
0.508	0.956	-5.930	0.0835	0.00000	0.0315	0.636	149.710	240512	dangdut
0.732	0.328	-7.72	0.0327	0	0.0864	0.55	112.102	165000	indonesian indie
0.618	0.776	-6.410	0.0282	0.00000	0.1270	0.842	133.986	300513	classic indonesian rock
0.503	0.859	-4.423	0.0455	0	0.0764	0.414	114.063	223613	indonesian pop
0.549	0.797	-6.606	0.0572	0.0035	0.0863	0.527	145.093	227187	indonesian folk

Pada tabel di atas menunjukkan nilai dari setiap audio analisis dari genre masing-masing musik yang sudah dihasilkan dari spotify. Melalui data tersebut selanjutnya akan ditransformasi ke dalam bentuk citra.

**Gambar 4 Transformasi data tabular ke bentuk citra**

Citra di atas yang selanjutnya masuk ke model untuk klasifikasi genre musik indonesia. Pada proses training data citra berjumlah 1160 dilatih sebanyak 10 epoch dengan size setiap

batch berjumlah 128. arsitektur model menggunakan DenseNet121 lalu ditambah dengan layer GlobalAveragePool2D kemudian ditambahkan 1 layer neural network dengan jumlah node 16 menggunakan fungsi aktivasi Relu dan output layer dengan fungsi aktivasi softmax. Model juga menggunakan learning rate optimizer fungsi Adamax dan fungsi loss sparse categorical crossentropy.



**Gambar 5 Akurasi dan Loss**

Setelah proses training sebanyak 10 epoch didapatkan hasil dari proses training dengan metrics accuracy dan metrics loss. Proses klasifikasi berhasil mendapatkan akurasi pada data latih mencapai nilai tertinggi 0.6713 dan loss terendah 0.8923 sedangkan hasil dari data uji akurasi tertinggi mencapai 0.3448 dan loss terendah 1.52.

## KESIMPULAN DAN SARAN

Berdasarkan hasil pelatihan dan evaluasi Implementasi SuperTML untuk klasifikasi Genre Musik Indonesia dari dataset yang terkumpul berjumlah 1160 yang diambil dari spotify. Menghasilkan performa yang tidak terlalu baik model dapat menghasilkan nilai akurasi tertinggi 0.6713 pada proses training dan 0.3448 pada proses uji. Namun dengan klasifikasi Genre musik Indonesia menggunakan citra terlihat memiliki potensi yang besar.

Dalam pengembangan kedepannya dapat ditambahkan tahapan preprocessing sebelum data tabular ditransformasikan ke bentuk citra serta dapat dilakukan perbaikan arsitektur model sehingga dapat menghasilkan performa yang lebih baik kedepannya.

## DAFTAR REFERENSI

- Arjaya, D. (2016). Dangdut dan Rezim Orde Baru: Wacana Nasionalisasi Musik Dangdut Tahun 1990-an. *Universitas Gajah Mada*, 12.
- Ayu, G., & Giri, V. M. (2017). KLASIFIKASI DAN RETRIEVAL MUSIK BERDASARKAN GENRE (SEBUAH STUDI PUSTAKA). In *Jurnal Ilmiah ILMU KOMPUTER Universitas Udayana: Vol. X* (Issue 1). [www.allmusic.com](http://www.allmusic.com)
- Lidinillah Alfath, N., Emanuela, O., & Alya Fatma, dan. (2022). *PENGARUH MUSIK POPULER DALAM MEMBANTU TINGKAT PEMAHAMAN PEMBELAJARAN MATEMATIKA TERHADAP SISWA SMA* (Vol. 2, Issue 1).
- Phung, V. H., & Rhee, E. J. (2019). A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. *Applied Sciences (Switzerland)*, 9(21). <https://doi.org/10.3390/app9214500>
- West, Jeremy; Ventura, Dan; Warnick, Sean (2007). "Spring Research Presentation: A Theoretical Foundation for Inductive Transfer". Brigham Young University, College of Physical and Mathematical Sciences. Archived from the original on 2007-08-01. Retrieved 2007-08-05.
- B. Sun et al., "SuperTML: Two-dimensional word embedding for the precognition on structured tabular data," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2019, pp. 1–9
- G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708