AN ENSEMBLE OF SIMPLE CONVOLUTIONAL NEURAL NETWORK MODELS FOR MNIST DIGIT RECOGNITION

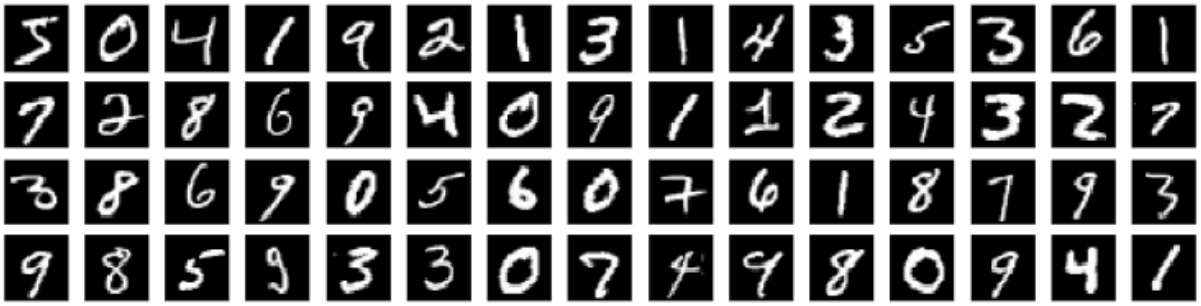
ABSTRAK

Akurasi yang sangat tinggi pada set tes MNIST dapat dicapai dengan menggunakan model jaringan saraf convolutional (CNN) sederhana. Menggunakan tiga model berbeda dengan 3×3, 5×5, dan 7×7 ukuran kernel di lapisan konvolusi. Setiap model terdiri dari satu set lapisan konvolusi diikuti oleh satu lapisan yang terhubung penuh. Setiap lapisan konvolusi menggunakan normalisasi batch dan aktivasi ReLU, sedangkan pooling tidak digunakan. Rotasi dan translasi digunakan untuk menambah data pelatihan, yang merupakan teknik yang sering digunakan di sebagian besar tugas klasifikasi gambar.

Kata Kunci Image Clasification MNIST

1. Introduction

Kumpulan data pengenalan digit tulisan tangan MNIST (Gambar 1) adalah salah satu kumpulan data paling dasar yang digunakan untuk menguji kinerja model jaringan saraf dan teknik pembelajaran. Menggunakan 60.000 gambar sebagai set pelatihan, akurasi 97% -98% dapat dengan mudah dicapai pada set uji 10.000 gambar, dengan metode pembelajaran seperti k-nearest tetangga (KNN), hutan acak, mesin vektor dukungan (SVM) dan model jaringan saraf sederhana. Jaringan saraf convolutional (CNN) meningkatkan akurasi ini hingga lebih dari 99% dengan kurang dari 100 gambar yang salah diklasifikasikan dalam set pengujian.



Gambar 1 Gambar dari MNIST training set.

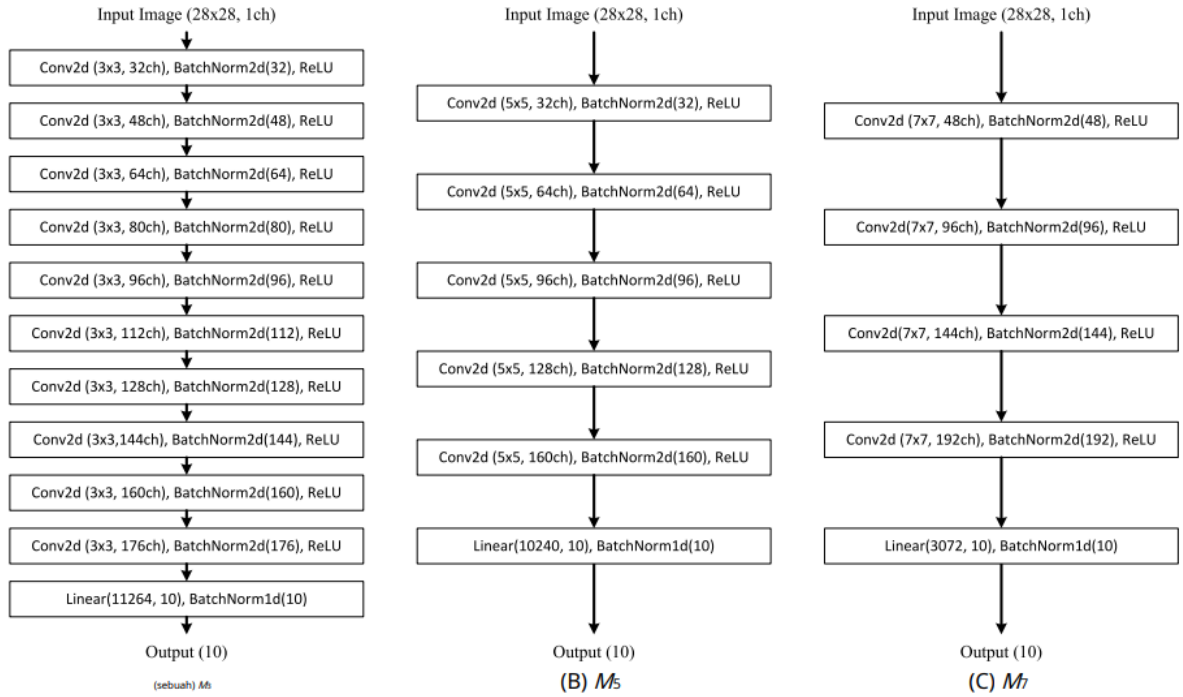
100 gambar terakhir lebih sulit untuk diklasifikasikan dengan benar. Untuk meningkatkan akurasi setelah 99%, membutuhkan model yang lebih kompleks, penyetelan hiperparameter yang cermat seperti kecepatan pembelajaran dan ukuran batch, teknik regularisasi seperti normalisasi *batch* dan *dropou*t, dan augmentasi data pelatihan. Ketiga model memiliki arsitektur yang serupa, tetapi memiliki ukuran kernel yang berbeda pada lapisan konvolusi. Eksperimen menunjukkan bahwa menggabungkan model dengan ukuran kernel yang berbeda mencapai akurasi yang lebih baik daripada menggabungkan model dengan ukuran kernel yang sama.

2. Netowrk Design dan Training

Di setiap lapisan konvolusi, konvolusi 2D dilakukan, diikuti oleh normalisasi batch 2D dan aktivasi ReLU. Penyatuan maksimum atau penyatuan rata-rata tidak digunakan setelah konvolusi. Sebagai gantinya, ukuran peta fitur berkurang setelah setiap konvolusi karena padding tidak digunakan. Misalnya, jika kita menggunakan 3×3 kernel, lebar dan tinggi gambar dikurangi dua setelah setiap lapisan konvolusi. Pendekatan serupa diambil di jaringan lain [6, 2]. Jumlah saluran ditingkatkan setelah setiap lapisan untuk memperhitungkan pengurangan ukuran peta fitur.

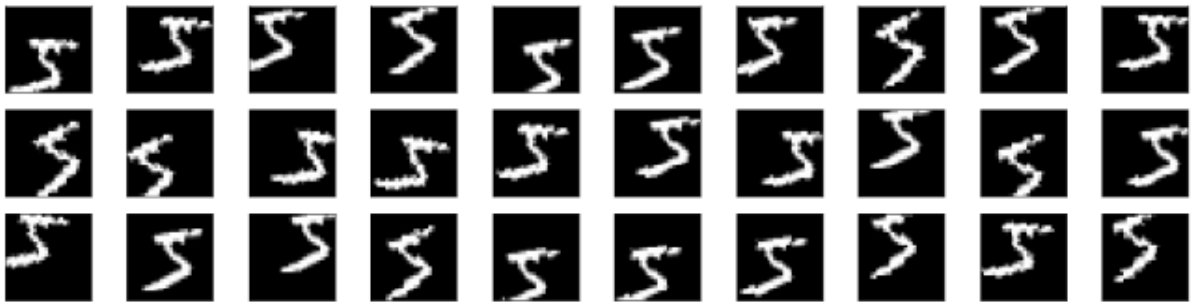
Menggunakan tiga jaringan berbeda dan menggabungkan hasil dari jaringan ini. Jaringan hanya berbeda dalam ukuran kernel dari lapisan konvolusi: 3×3, 5×5, dan 7×7. Karena ukuran kernel yang berbeda menyebabkan pengurangan ukuran yang berbeda dalam peta fitur, jumlah lapisan berbeda untuk setiap

jaringan. Jaringan pertama,M3, menggunakan 10 lapisan konvolusi dengan 16(Saya+ 1)saluran diSayalapisan konvolusi. Peta fitur menjadi 8×8 dengan 176 saluran setelah lapisan ke-10. Jaringan kedua,M5, menggunakan 5 lapisan konvolusi dengan32Sayasaluran diSayalapisan konvolusi. Peta fitur menjadi 8×8 dengan 160 saluran setelah lapisan ke-5. Jaringan ketiga,M7, menggunakan 4 lapisan konvolusi dengan48Sayasaluran diSayalapisan konvolusi. Peta fitur menjadi 4×4 dengan 192 saluran setelah lapisan ke-4. Struktur ketiga jaringan tersebut ditunjukkan pada Gambar 2.



Gambar 2 Network models digunakan untuk MNIST digit classification.

Untuk terjemahan acak, gambar digeser secara acak secara horizontal dan vertikal, hingga 20% dari ukuran gambar di setiap arah. Untuk rotasi acak, gambar diputar hingga 20 derajat baik searah jarum jam atau berlawanan arah jarum jam. Jumlah transformasi bervariasi untuk setiap gambar dan setiap zaman, sehingga jaringan dapat melihat berbagai versi gambar dalam set pelatihan (Gambar 3).



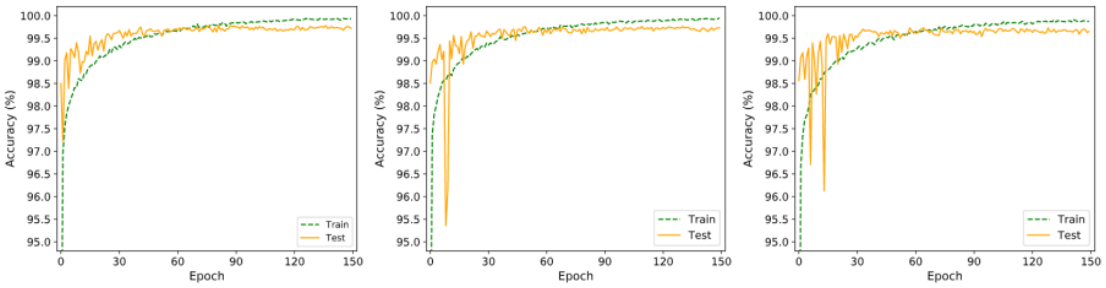
Gambar 3 Random Translation dan Random Rotation diterapkan pada training image.

menggunakan pengoptimal Adam dengan fungsi kehilangan lintas entropi. Tingkat pembelajaran dimulai pada 0,001, dan secara eksponensial meluruh dengan faktor peluruhan γ.=0,98. Ukuran batch adalah 120, dan 500 pembaruan parameter dan menggunakan bobot rata-rata bergerak eksponensial untuk evaluasi, yang dapat mengarah pada generalisasi yang lebih baik. Peluruhan eksponensial yang digunakan untuk menghitung rata-rata bergerak adalah 0,999.

3. Experiment

3.1 Results for Individual Networks and Ensembles

Dalam hal akurasi pengujian, jaringan dengan kernel yang lebih besar menunjukkan beberapa ketidakstabilan pada zaman awal, tetapi pola semua jaringan menjadi serupa setelah 50 zaman. Tabel 1 menunjukkan akurasi minimum, rata-rata, maksimum dari 30 jaringan antara 50 dan 150 zaman, dalam kisaran kepercayaan 95%. Akurasi dariM3 sedikit lebih tinggi diikuti olehM5 dan M7, tetapi perbedaannya tidak terlalu signifikan (kurang dari 0,02%). Antara 50 dan 150 zaman dari 30 jaringan, akurasi pengujian tertinggi diamati dariM3, M5, M7 masing-masing adalah 99,82, 99,80, dan 99,79.



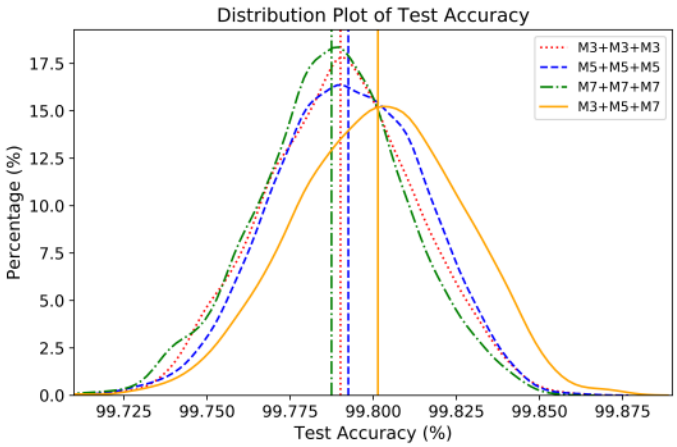
(a) M_3 (b) M_5 (c) M_7

Gambar 4 Akurasi train dan test akurasi dari M3, M5, M7 selama Training.

Table 1 Test akurasi dari Networks yang diukur antara 50 epoch dan 150 epoch dalam train.

model	test accuracy			
	min	avg	max	best
M_3	99.5930 ± 0.0136	99.6949 ± 0.0058	99.7667 ± 0.0084	99.82
M_5	99.5863 ± 0.0115	99.6835 ± 0.0074	99.7583 ± 0.0081	99.80
M_7	99.5470 ± 0.0288	99.6711 ± 0.0089	99.7450 ± 0.0093	99.79

Hasil akhir diperoleh dengan menggunakan suara terbanyak. Artinya, jika dua jaringan setuju bahwa gambar milik kelas tertentu, kelas itu dipilih. Jika tiga jaringan memilih kelas yang berbeda, satu kelas dipilih secara acak di antara ketiganya. Untuk setiap strategi, kami menguji 1000 jaringan ansambel dan memplot histogram untuk akurasi pengujian. UntukM3, M5, dan M7, akurasi pengujian yang lebih tinggi dapat dicapai dengan menggabungkan hasil dari tiga jaringan. Dapat diamati bahwa sementara akurasi uji rata-rata dari metode ensemble homogen serupa, metode ensemble di mana satu jaringan dipilih dari setiap jenis jaringan mencapai akurasi yang lebih tinggi.



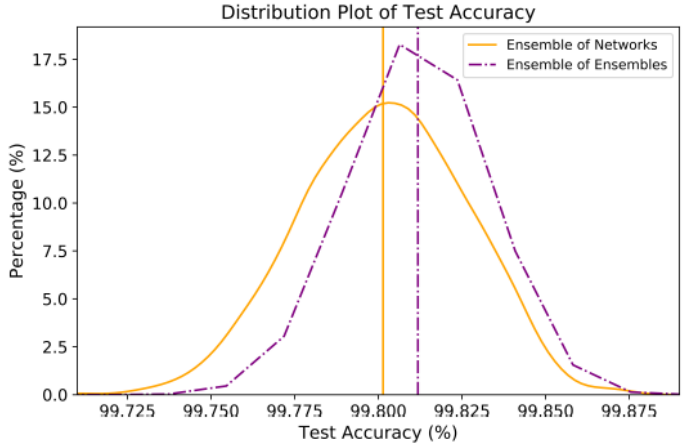
Gambar 5 Distribusi akurasi uji untuk jaringan ensemble homogen dan heterogen

Secara acak memilih 3 jaringan dariM3 dan menggabungkan hasil mereka menggunakan suara mayoritas. Demikian pula, kami menggabungkan hasil dari tiga jaringan untukM5 dan M7. Setelah itu, kami menggunakan suara mayoritas untuk tiga jaringan ansambel. Gambar 6 menunjukkan distribusi akurasi pengujian untuk 1000 ensemble jaringan individu (M3+M5+M7) dan 1000 ansambel jaringan ansambel ((M3+M3+M3)+(M5+M5+M5)+(M7+M7+M7)). Grafik menunjukkan bahwa menggunakan ensemble jaringan ensemble meningkatkan akurasi tes rata-rata. Tabel 3 menunjukkan rentang kepercayaan 95% dan akurasi terbaik yang diamati untuk ensemble individu dan ensemble jaringan ensemble.

Table 2 confidence range 95% akurasi uji untuk jaringan ensemble homogen dan heterogen.

configuration	test accuracy	
	95% confidence range	best accuracy
$M_3 + M_3 + M_3$	99.7901 ± 0.0014	99.86
$M_5 + M_5 + M_5$	99.7925 ± 0.0014	99.86
$M_7 + M_7 + M_7$	99.7874 ± 0.0014	99.85
$M_3 + M_5 + M_7$	$\mathbf{99.8014 \pm 0.0015}$	**99.87**

Selain pemilihan acak, kami juga menunjukkan kasus terbaik untuk melihat akurasi terbaik yang dapat kami capai. Untuk kasus terbaik, kami memilih 10 jaringan ansambel homogen dariM3, M5, dan M7 yang menunjukkan akurasi tes terbaik. Kemudian, kami memilih satu jaringan dari setiap jenis dan menggabungkan hasilnya. Akurasi terbaik yang dicapai adalah 99,91%.



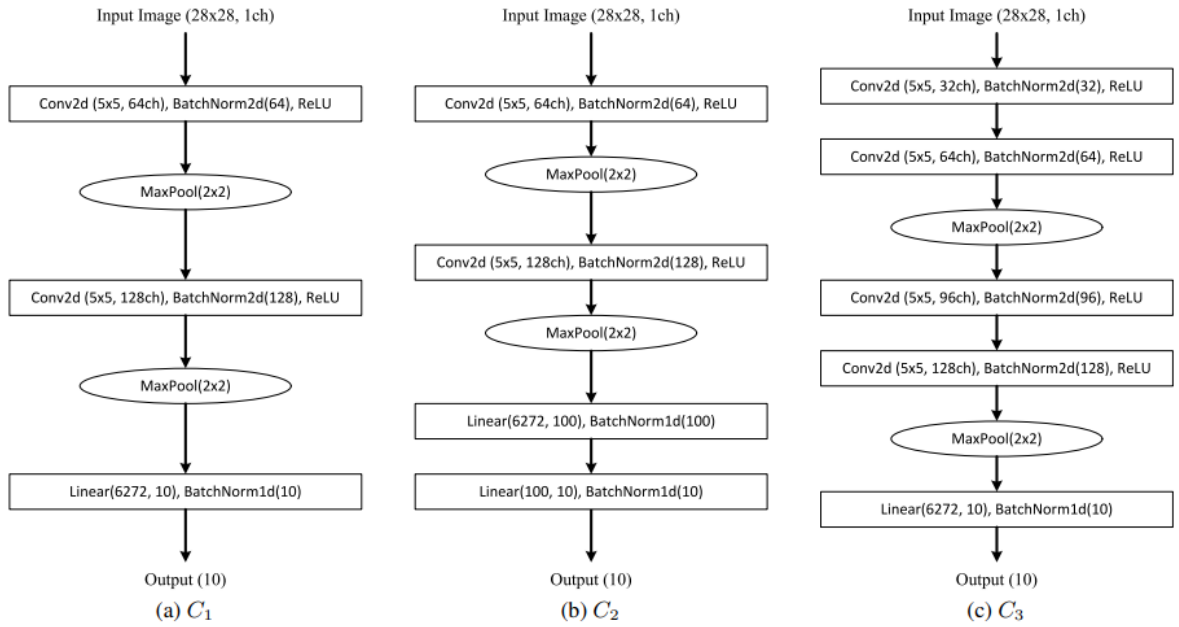
Gambar 6 Distribusi akurasi uji untuk ensemble jaringan individu dan ensemble jaringan ensemble.

Table 3 95% confidence range and best-case untuk ensemble jaringan individual dan ensemble jaringan

configuration	test accuracy	
	95% confidence range	best accuracy
ensemble of individual networks	99.8014 ± 0.0015	99.87
ensemble of ensembles (random)	99.8118 ± 0.0002	99.89
ensemble of ensembles (best)	99.8646 ± 0.0008	99.91

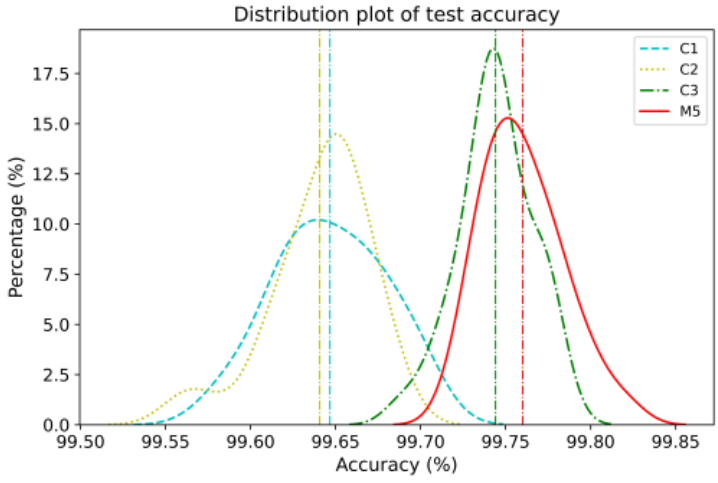
3.2 Dampak Network architecture

Model CNN yang umum digunakan terdiri dari satu set lapisan konvolusi di mana setiap lapisan konvolusi diikuti oleh lapisan penyatuan, dan satu atau beberapa lapisan yang terhubung penuh pada akhirnya. Beberapa jaringan memiliki dua lapisan konvolusi sebelum lapisan penyatuan.



Gambar 7 Struktur CNN yang umum digunakan dengan max pooling.

Akurasi tes rata-rata dariC3 dan M5 lebih baik dariC1dan C2, yang berarti menggunakan lebih banyak lapisan konvolusi dapat menghasilkan pembelajaran fitur yang lebih baik. Memiliki lebih banyak lapisan yang terhubung sepenuhnya pada akhirnya tidak membantu, seperti yang dapat dilihat dari keakuratanC1dan C2. Di antaraC3 dan M5, M5 mencapai akurasi yang lebih tinggi secara umum, dan juga dapat mencapai akurasi yang lebih tinggi dalam kasus terbaik.



Gambar 8 Distribusi akurasi pengujian untuk jaringan dengan arsitektur berbeda.

Table 4 akurasi pengujian untuk jaringan dengan arsitektur yang berbeda.

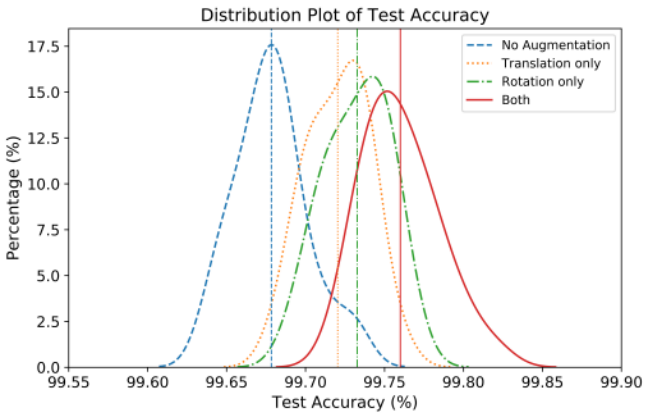
model	akurasi tes
C_1	99.6466 ± 0,0121
C_2	99.6406 ± 0,0108
C_3	99.7440 ± 0,0080
M_5	99.7600 ± 0,0089

3.3 Dampak augmentasi data

Augmentasi data adalah teknik untuk meningkatkan keragaman data pelatihan tanpa benar-benar mengumpulkan data dan melabelinya. Ini adalah teknik penting untuk pembelajaran terawasi di mana kumpulan data besar diperlukan untuk model jaringan untuk mencapai kinerja tinggi [14, 15, 16, 17, 18]. Untuk kumpulan data MNIST, menerapkan rotasi acak memiliki kontribusi yang sedikit lebih tinggi daripada terjemahan acak, tetapi kedua skema diperlukan untuk mencapai akurasi terbaik. Tabel 5 menunjukkan kisaran kepercayaan 95% dari akurasi tes untuk empat strategi augmentasi.

Table 5 Confidence range 95% dari akurasi pengujian untuk jaringan yang dilatih dengan skema augmentasi yang berbeda

skema augmentasi		akurasi tes
terjemahan	rotasi	
7	7	99.6783 ± 0,0086
3	7	99.7203 ± 0,0074
7	3	99.7327 ± 0,0077
3	3	99.7600 ± 0,0089

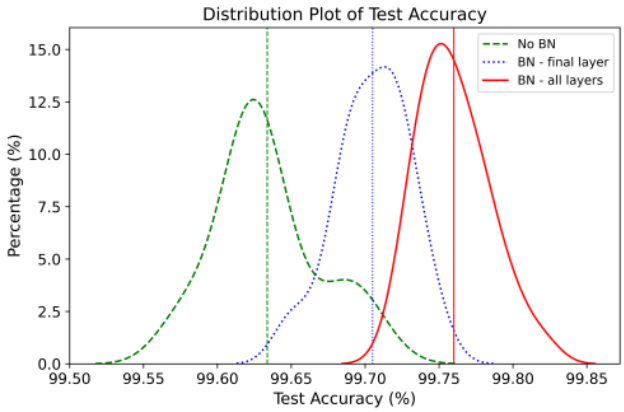


Gambar 9 Distribusi akurasi pengujian untuk jaringan yang dilatih dengan skema augmentasi yang berbeda.

3.4 Dampak normalisasi batch

Normalisasi batch adalah teknik terkenal untuk meningkatkan kinerja jaringan serta stabilitas dan kecepatan pelatihan [19]. Telah dilaporkan bahwa sebagian besar model jaringan saraf mendapat manfaat dari penggunaan normalisasi batch [20, 21]. Dampak normalisasi batch pada kinerja model jaringanM5. Kami membandingkan tiga konfigurasi: model pertama tidak menggunakan normalisasi

batch sama sekali, model kedua menggunakan normalisasi batch hanya pada lapisan yang terhubung penuh, dan model ketiga menggunakan normalisasi batch di semua lapisan.



Gambar 10 Distribusi akurasi pengujian untuk jaringan yang dilatih dengan skema normalisasi batch yang berbeda.

Table 6 Rentang kepercayaan 95% dari akurasi pengujian untuk jaringan yang dilatih dengan skema normalisasi batch yang berbeda.

konfigurasi	akurasi tes
tidak ada normalisasi batch	99.6337 ± 0,0131
normalisasi batch di lapisan akhir	99.7050 ± 0,0092
normalisasi batch di semua lapisan	99.7600 ± 0,0089

4. Kesimpulan

Kumpulan data digit tulisan tangan MNIST sering digunakan sebagai kumpulan data tingkat awal untuk pelatihan dan pengujian jaringan saraf. Meskipun mencapai akurasi 99% pada set pengujian agak mudah, mengklasifikasikan 1% gambar terakhir dengan benar merupakan tantangan. Orang-orang telah mencoba banyak model dan teknik jaringan yang berbeda untuk meningkatkan akurasi pengujian, dan akurasi terbaik yang dilaporkan mencapai sekitar 99,8%. Dalam makalah ini kami menunjukkan bahwa model CNN sederhana dengan normalisasi batch dan augmentasi data dapat mencapai akurasi terbaik. Menggunakan ansambel model jaringan homogen dan heterogen dapat meningkatkan kinerja, akurasi uji hingga 99,91% yang merupakan salah satu kinerja canggih. Studi dengan berbagai konfigurasi yang berbeda menunjukkan bahwa kinerja tinggi tidak dicapai dengan teknik tunggal atau arsitektur model,