تشخیص بیماریهای رایج گل کلم با استفاده از پردازش تصویر و یادگیری عمیق راضیه پوردربانی ۱*, سجاد سبزی۲

چکیده

گل کلم یک سبزی بسیار سالم است که منبع مهمی از مواد مغذی است. گل کلم به طور طبیعی سرشار از فیبر و ویتامین B است. آنتی اکسیدان ها و مواد مغذی گیاهی را فراهم می کند که می تواند در برابر سرطان محافظت کند. همچنین حاوی فیبر برای کاهش وزن و هضم غذا، کولین که برای یادگیری و حافظه ضروری است و بسیاری از مواد مغذی مهم دیگر است. عواملی همچون کمبود مواد مغذی، شرایط آب و هوایی و بیماریها باعث میشود که رشد گل کلم همراه با مشکلاتی باشد. گل کلم حاوی بیماری های پاتولوژی گیاهی مختلفی است؛ اما روش سنتی برای شناسایی کلم های بیمار و جداسازی آنها خسته کننده و وقت گیر است بنابراین می توان با استفاده از تکنیک تصویربرداری و شبکه عصبی مصنوعی در کوتاهترین زمان ممکن آنها را شناسایی کرد. هدف از این تحقیق، طیقه بندی گل کلم به ۴ گروه سالم، آلوده به کپک پودری، پوسیدگی سیاه و پوسیدگی نرم باکتریایی با استفاده از تکنیک پردازش تصویر و یادگیری عمیق LeNet می باشد. ابتدا مجموع ۶۵۵ تصویر رنگی شامل ۴ کلاس مذکور تهیه شد. ۷۰ درصد داده ها برای آموزش مدل در نظر گرفته شد. نتایج نشان داد که مدل توانست گل کلم های سالم، گل کلم های آلوده به پوسیدگی سیاه و کپک پودری تواستند با دقت ۷۰۰ درصد شناسایی شوند. گل کلم های آلوده به پوسیدگی نرم باکتریایی تواستند با دقت ۲۰۰ درصد شناسایی شوند. گل کلم های آلوده به پوسیدگی نرم باکتریایی تواستند با دقت ۲۰۰ درصد شناسایی شوند. گل کلم های آلوده به پوسیدگی نرم باکتریایی تواستند با دقت ۲۰۰ درصد شناسایی شوند. گل کلم های آلوده به پوسیدگی نرم باکتریایی تواستند با دقت ۲۰۹ درصد شناسایی شوند.

كلمات كليدي

"گل کلم"، "پردازش تصویر"، "بیماری"، "یادگیری عمیق".

١- مقدمه

گل کلم یک منبع عالی از ویتامین C ،فولات، ویتامین K ، ویتامین های پیچیده B و ویتامین E است. گل کلم همچنین، حاوی ماده معدنی حیاتی، مانند: کلسیم، منیزیم، فسفر، پتاسیم، سدیم و آهن است که بدون اضافه کردن کلسترول، برای بدن مضر نمی باشند. همچنین یک منبع خوبی از پروتئین و چربی های اشباع نشده و اسیدهای چرب امگا ۳ است و فیبر رژیمی را فراهم می کند و مقادیر کمتری از قندهای طبیعی را در مقایسه با سایر اعضای خانواده گیاه شناسی مانند برو کلی دارد . عواملی همچون کمبود مواد مغذی، شرایط آب و هوایی و بیماری ها باعث می شود که رشد گل کلم همراه با مشکلاتی باشد. برخی از بیماری های شناخته شده گل کلم عبارتند از: پوسیدگی سیاه ، کپک پودری ، پوسیدگی نرم باکتریایی و ساق سیاه . پوسیدگی سیاه توسط قارچی به نام Alternaria brassicae ایجاد می شود. ابتدا لکه های حلقه ای سیاه رنگ روی برگ های پایینی گل کلم مشاهده می شود. سپس در مرحله پیشرفته، برگ ها زرد شده و در نهایت می ریزند. کپک پودری نیز توسط قارچی به نام Peronospora parasitica یجاد می شود که در قسمت بالایی برگ به صورت لکه های کوچک زرد رنگ دیده می شود و در نهایت قه وه ای می شود. ساق سياه توسط Phoma lingam ايجاد مي شود. اين قارچ لكه های زرد تا قهوه ای با مرکز خاکستری روی گیاه ایجاد می کند و می تواند نهال ها را از بین ببرد. بیماری های پاتولوژی گیاهی دارای انـواع مختلفی هستند. برخی از آنها ممکن است با چشم غیرمسلح قابل مشاهده نباشند، بنابراین میکروسکوپ های قدرتمندی مورد نیاز است. تعدادی از آنها را می توان با استفاده از تکنیک تصویربرداری چند طیفی و فراطیفی شناسایی کرد. اگر علائم بیماری قابل مشاهده باشد، پردازش تصویر یکی از گزینه های امیدوارکننده است (عبداله و

همکاران، ۲۰۰۷۸؛ ویرندرا و همکاران، ۲۰۲۲). در واقع، در سیستم تشخیص بیماری، دادههای اموزش کافی باید ارائه شود. در مرحله بعد، مقایسه مداوم بر روی داده های نمونه با داده های آموزشی انجام شد و خروجی بر اساس حداقل خطای بین آن دو داده ارائه شـد (بادیگر و همکاران ۲۰۲۲). محققان مختلف تکنیک های پردازش تصویر را بر اساس ویژگی های مختلف مانند رنگ، بافت و شکل انجام داده اند. برخی عملکرد خوبی به ویژه در شرایط کنترل شده نشان داده اند، با این حال، تشخیص دقیق در شرایط کنترل نشده چالش برانگیز باقی می ماند . روش سنتی برای شناسایی هر یک ازمحصولات کشاورزی، بازرسی بینایی است که کار بسیار خسته کننده و وقت گیر است یک سیستم بینایی کامپیوتری یا سیستم بینایی ماشین شامل زمینه های مختلف از قبیل عملیات پزشکی، نظارت بر ترافیک، بازرسی محصولات صنعتی و کشاورزی و صنایع غذایی مانند طبقه بندی، برداشت و ارزیابی اتوماتیک و شناسایی ارقام مختلف دانه به عنوان ارزیابی غیرمخرب میباشد (هنگ و همکاران، ۲۰۱۵). برخی از محققانی که از سیستم های بینایی ماشین برای طبقه بندی دانه های مختلف استفاده می کنند (آژاکس و حسین، ۲۰۱۵؛ ولاسکو و فـدیف، ۲۰۱۷؛ نایـک و همکاران، ۲۰۱۷؛ فوزی ۲۰۱۸). راج و همکاران (۲۰۲۲) یک طبقهبندی بر اساس بینایی کامپیوتری و شبکه عصبی کانولوشن کانتور فعال عمیق برای تقسیمبندی بیماری محصول بر اساس لکه برگ سرکوسپورا، سوختگی باکتریایی، کپک پودری و زنگ ارائه کردند. آنادهکریشنان و جیاسشختی در سال ۲۰۲۲، یک شبکه عصبی پیچیده عمیق برای تشخیص بیماری برگ گوجه فرنگی مبتنی بر تصویر ایجاد کردند. آنها از مجموعه داده دارای ۱۸۱۶۰ تصویر استفاده کردند که در آن ۶۰ درصد از تصاویر برای آموزش و ۴۰ درصد برای تست استفاده شده است. در نهایت، آنها دقت ۹۸٬۴۰٪ را برای مجموعه تست به

دست آوردند. ونگچای و همکاران (۲۰۲۲) یک سیستم یادگیری عمیق نظارتی را برای پیش بینی بیماری محصول پیشنهاد کرد. نتایج ارزیابی پارامترهای سیستم به ترتیب ۹۶، ۸۹، ۸۹، ۷۵ درصد و ۶۶ درصد برای دقت، دقت، ویژگی، امتیاز F و AUC بود. ژو و همکاران (۲۰۲۲) مطالعه ای بر روی دسته بندی گل کلم بر اساس تصویربرداری تشدید مغناطیسی انجام دادند. آموزش گل کلم سالم و تحت استرس با طبقه بندى باينرىPLSDA ، QDA ، LDAو CNN انجام شد. نتایج نرخ طبقه بندی باینری و امتیاز F را تا ۹۵ درصد نشان داد. حمودا و همکاران (۲۰۱۷) الگوریتمی بر اساس فضای رنگی HSV برای تشخیص گل کلم از علف های هرز و خاک پیشنهاد کرد. ناحیه مورد علاقه (ROI) با فیلتر کردن هر یک از کانال های تعیین شد. نتایج به دست آمده با تفسیر دستی مقایسه شد. عملکرد الگوریتم با حساسیت و دقت به ترتیب با مقدار ۹۸٬۹۱٪ و ۹۹٬۰۴٪ ارزیابی شد. با توجه به خواص فوق العاده غذایی و درمانی گل کلم، حفظ سلامت محصول و جداسازی گیاهان بیمار از اهمیت ویژه ای برخوردار است. شایان ذکر است که از این گیاهان بیمار می توان برای تهیه کمپوست استفاده کرد (کوماری و همکاران، ۲۰۲۰). هدف از این تحقیق، طیقه بندی گل کلم به ۴ گروه سالم، آلوده به کپک پودری، پوسیدگی سیاه و پوسیدگی نرم باکتریایی با استفاده از تکنیک پردازش تصویر و یادگیری عمیق می باشد.

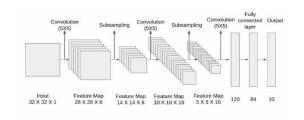
٢- روش انجام تحقيق

ابتدا مجموع ۶۵۵ تصویر شامل ۴ کلاس سالم، آلوده به کپک پودری، آلوده به پوسیدگی سیاه و آلوده به پوسیدگی نرم باکتریایی تهیه شد. از این تعداد تصاویر، ۷۰ درصد برای داده آموزش، ۱۰ درصد برای داده اعتبارسنجی و ۲۰ درصد برای داده تست اختصاص داده شد. نمونه ای از تصاویر مربوط به هر کلاس در شکل ۱ اورده شده است.



شکل ۱. نمونه از تصاویر مربوط به کلاس های سالم، اَلوده به کپک پودری، اَلوده به پوسیدگی سیاه و اَلوده به پوسیدگی نرم باکتریایی در گل کلم

برای تحلیل تصاویر از شبکه عصبی کانولوشن دوبعدی LeNet برای تحلیل استفاده شد. این مدل در سال ۱۹۹۸ اولین بار در مقالهای با نام Gradient-Based Learning Applied to Document رونمایی شد. ساختار مدل لی نت در شکل ۲ آمده



شكل ٢. ساختار مدل شبكه عصبى كانولوشن LeNet

به عنوان مثال اگر دو طبقه وجود داشته باشد ماتریس درهم ریختگی ۲ ×۲ می باشد. با استفاده از این ماتریس معیارهای مختلف قابل تعریف و اعمال می باشند که در ادامه تعدادی از این معیارها بررسی خواهند شد (پوردربانی و رضایی، ۲۰۱۱؛ سبزی و همکاران، ۲۰۲۱). حساسیت یا بازخوانی:

چند درصد از نمونههای درست، بهدرستی تشخیص داده شدهاند (معادله ۱).

$$Recall = \frac{TP}{TP + FN} \times 100 \tag{1}$$

دقت:

درصد کل پاسخهای صحیح سیستم (معادله ۲)

Accuracy =
$$\frac{TP + TN}{TP + FN + FP + TN} \times 100$$
 (Y)

صحت:

چند درصد از خروجیهای درست تشخیص داده شده، واقعاً درست

است (معادله ۳).

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{(7)}$$

معيار F:

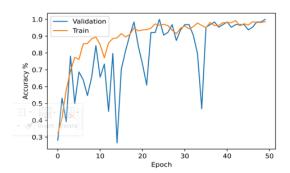
میانگین وزن دار هارمونیک Recall و Precision (معادله ۴)

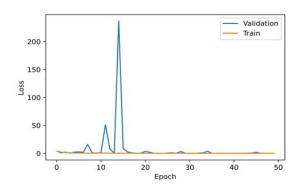
$$F_{\text{measure}} = \frac{2 \times \text{Recall} \times Precision}{\text{Recall} + Precision}$$
 (*)

که TP تعداد نمونههای هر کلاس که درست طبقهبندی شدهاند. TP برابر است با تعداد نمونههای روی قطر اصلی ماتریس کنفیوژن منهای تعداد نمونههایی که بهدرستی در کلاس موردنظر طبقهبندی شدهاند. FN بهصورت مجموع نمونههای افقی کلاس مورد بررسی منهای تعداد نمونههایی که بهدرستی در کلاس موردنظر طبقهبندی شدهاند تعریف می گردد. و FP مجموع نمونههای عمودی کلاس مورد بررسی منهای تعداد نمونههایی که بهدرستی در کلاس مورد نظر طبقهبندی شدهاند، می باشد.

٣- نتايج

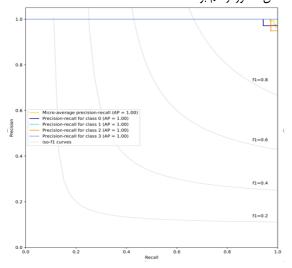
شکل ۳ نمودار دقت را برای ۵۰ اپوک (تکرار) به نمایش میگدارد. همونطور که از نمودار پیداست، مدل توانسته که بعد از اپوک پانزدهم با دقت بالای ۹۰٪ اَموزش ببیند. بعد از اپوک سی و پنجم، نمودار اعتبارسنجی و اَموزش تقریبا باهم منطبق شدند. نمودار شکل ۴ نیز موید این مطلب هست که بعد از اپوک ۳۵م، اَموزش با دقت بسیار بالا و حداقل خطا صورت پذیرفت و اعتبارسنجی گشت.



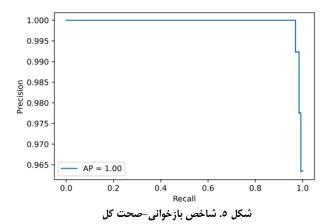


شکل ۳. نمودار دقت به ازای اپوک های مختلف برای داده آموزش و اعتبار

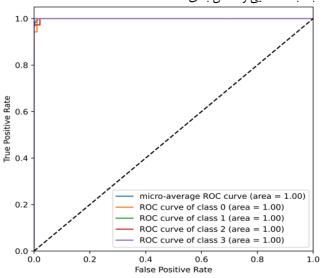
در شکل ۴، نتایج نمودار بازخوانی-صحت برای تک تک کلاسهای تحقیق حاضر نمایش داده می شود. شاخص ۱۰۰ درصدی بازخوانی-صحت برای هر کلاس، نشان از موفقیت طبقه بند در شناسایی ۴ کلاس مذکور از هم بوده است.



شکل ٤. شاخص بازخوانی –صحت به ازای هر کلاس مورد بررسی؛ کلاس + تا ٤ به ترتیب اشاره به سالم، آلوده به کپک پودری، پوسیدگی سیاه و پوسیدگی نرم باکتریایی دارد



شکل ۶ منحنی عملیاتی دریافتکننده (ROC) را برای هر کلاس نشان می دهد. با توجه به نزدیک بودن تمام نمودارها به خط قایم و عدد یک، می توان اظهار داشت که تمام کلاسها با موفقیت کامل توسط طبقه بند شناسایی و کلاس بندی شدند.



شکل ٦. منحنی عملیاتی دریافت کننده (ROC) را برای هر کلاس. کلاس + تا ٤ به ترتیب اشاره به سالم، اَلوده به کپک پودری، پوسیدگی سیاه و پوسیدگی نرم باکتریایی دارد

عملکرد طبقه بند در قالب ماتریس اغتشاش و پارامترهای دقت، صحت، بازخوانی و شاخص اف در جدول ۱ و ۲ آورده شده است.

جدول ۱. عملكرد طبقه بند در قالب ماتریس اغتشاش

جدون ۱۰ حسوره طبعه بند در عاب ما فریس احساس						
کپک پودری	پوسیدگی نرم باکتریایی	پوسید <i>گی</i> سیاه	سالم			
•	•	•	۳۵	سالم		
•	٠	۲٠	٠	پوسیدگ <i>ی</i> سیاه		
•	775	٠	١	پوسیدگی نرم باکتریایی		
۴٠	١	١	١	کپک پودری		

تهیه کمپوست استفاده کرد. در این تحقیق، طیقه بندی گل کلم به ۴ گروه سالم، آلوده به کپک پودری، پوسیدگی سیاه و پوسیدگی نرم باکتریایی با استفاده از تکنیک پردازش تصویر و یادگیری عمیق LeNet انجام شد که اهم نتایج به قرار زیر است:

- مدل توانست که با دقت بالای ۹۹٪ و حداقل خطا آموزش ببیند و اعتبارسنجی گردد.
- گل کلم های سالم، گل کلم های آلوده به پوسیدگی سیاه و کیک یودری تواستند با دقت ۱۰۰ درصد شناسایی شوند
- گل کلم های آلوده به پوسیدگی نرم باکتریایی تواستند با دقت ۹۹ درصد شناسایی شوند

جدول ۲. عملکرد طبقه بند در قالب پارامترهای دقت، صحت، بازخوانی و شاخص اف

شاخص اف	بازخواني	صحت	دقت	
(%)	(%)	(%)	(%)	
1	1	1	1	سالم
1	1	1	1	پوسیدگی سیاه
૧૧	૧૧	٩٨	૧૧	پوسی <i>دگی</i> نرم باکتریای <i>ی</i>
1	1	1	1	کپک پودری

٤- نتيجه گيري

روش سنتی برای شناسایی هر یک ازمحصولات کشاورزی، بازرسی بصری است که کار بسیار خسته کننده و وقت گیر است. با توجه به خواص فوق العاده غذایی و درمانی گل کلم، جداسازی گیاهان بیمار از اهمیت ویژه ای برخوردار است؛ بویژه اینکه گیاهان بیمار می توان برای

منابع

- Abdullah, N. E., Rahim, A. A., Hashim, H., & Kamal, M. M. (2007, December). Classification of rubber tree leaf diseases using multilayer perceptron neural network. In 2007 5th student conference on research and development (pp. 1-6). IEEE.
- Ahmad, I. S., Reid, J. F., Paulsen, M. R., & Sinclair, J. B. (1999). Color classifier for symptomatic soybean seeds using image processing. Plant disease, 83(4), 320-327.
- Anandhakrishnan, T., & Jaisakthi, S. M. (2022). Deep Convolutional Neural Networks for image based tomato leaf disease detection. Sustainable Chemistry and Pharmacy, 30, 100793.
- Badiger, M., Kumara, V., Shetty, S. C., & Poojary, S. (2022). Leaf and skin disease detection using image processing. Global Transitions Proceedings, 3(1), 272-278.
- Kumari, N., Sharma, A., Devi, M., Zargar, A., Kumar, S., Thakur, U., ... & Giri, A. (2020).
 Compost from the food waste for organic production of cabbage, cauliflower, and radish under sub-tropical conditions. International journal of recycling organic waste in agriculture, 9(4), 367-383.
- Pourdarbani, R., & Rezaei, B. (2011). Automatic detection of greenhouse plants pests by image analysis. Tarım Makinaları Bilimi Dergisi, 7(2), 171-174.
- Raj, N., Perumal, S., Singla, S., Sharma, G. K., Qamar, S., & Chakkaravarthy, A. P. (2022). Computer aided agriculture development for crop disease detection by segmentation and classification using deep learning architectures. Computers and Electrical Engineering, 103, 108357.
- Sabzi, S., Pourdarbani, R., Rohban, M. H., García-Mateos, G., Paliwal, J., & Molina-Martínez, J. M. (2021). Early detection of excess nitrogen consumption in cucumber plants using hyperspectral imaging based on hybrid neural networks and the imperialist competitive algorithm. Agronomy, 11(3), 575.
- Veerendra, G., Swaroop, R., Dattu, D. S., Jyothi, C. A., & Singh, M. K. (2022). Detecting plant diseases, quantifying and classifying digital image processing techniques. Materials Today: Proceedings, 51, 837-841.
- Wongchai, A., rao Jenjeti, D., Priyadarsini, A. I., Deb, N., Bhardwaj, A., & Tomar, P. (2022).
 Farm monitoring and disease prediction by classification based on deep learning architectures in sustainable agriculture. Ecological Modelling, 474, 110167.

Diagnosis of common cauliflower diseases using image processing and deep learning

Razieh Pourdarbani^{1*}; Sajad Sabzi²

*1- Associate Professor, Faculty of Agriculture, University of Mohaghegh Ardabili, Ardabil, Iran 2 - Faculty of Computer, Technical University of Sharif, Tehran, Iran *Email Address: r pourdarbani@uma.ac.ir

Abstract

Cauliflower is a very healthy vegetable that is an important source of nutrients that is naturally rich in fiber and B vitamins. It provides antioxidants and phytonutrients that can protect against cancer. It also contains fiber for weight loss and digestion, choline, which is essential for learning and memory, and many other important nutrients. Factors such as lack of nutrients, weather conditions and diseases cause the growth of cauliflower to be accompanied by problems. Cauliflower contains various plant pathology diseases; but the traditional method for identifying diseased cabbages and separating them is tedious and time-consuming, so they can be identified in the shortest possible time using imaging techniques and artificial neural networks. The purpose of this research is to categorize cauliflower into 4 healthy groups, infected with powdery mildew, black rot and bacterial soft rot using LeNet image processing and deep learning techniques. First, a total of 655 color images including the mentioned 4 classes were prepared. 70% of the data was considered for model training. The results showed that the model was able to identify healthy cauliflowers, cauliflowers infected with black rot and powdery mildew with 100% accuracy. Cauliflower infected with bacterial soft rot could be identified with 99% accuracy.

Introduction

Factors such as lack of nutrients, weather conditions and diseases cause the growth of cauliflower to be accompanied by problems. Some of the known diseases of cauliflower are: black rot, powdery mildew, bacterial soft rot and phoma Stem Kancer. Plant pathology diseases have different types. Some of them may not be visible to the naked eye, so powerful microscopes are needed. Some of them can be identified using multispectral and hyperspectral imaging techniques. If disease symptoms are visible, image processing is a promising option. The traditional method to identify any agricultural product is visual inspection, which is a very tedious and time-consuming task. A computer vision system or machine vision system includes various fields such as medical operations, traffic monitoring, inspection of industrial and agricultural products, and food industries like classification, harvesting and automatic evaluation and identification of different grain varieties as non-destructive evaluation. Some researchers who use machine vision systems to classify different seeds. Zhu et al. (2022) conducted a study on classification of cauliflower based on magnetic resonance imaging. The training of healthy and stressed cauliflower was done with LDA, QDA, PLSDA and CNN binary classification. The results showed binary classification rate and F score up to 95%. Hamouda et al. (2017) proposed an algorithm based on HSV color space to distinguish cauliflower from weeds and soil. A region of interest (ROI) was determined by filtering each of the HSV channels. The obtained results were compared with manual interpretation. Algorithm performance was evaluated with sensitivity and accuracy of 98.91% and 99.04%, respectively. Considering the extraordinary nutritional and therapeutic properties of cauliflower, maintaining the health of the product and separating diseased plants is of particular importance. It is worth mentioning that these infected plants can be used to prepare compost. The purpose of this research is to categorize cauliflower into 4 healthy groups, infected with powdery mildew, black rot and bacterial soft rot using image processing and deep learning techniques.

Methodology

First, a total of 655 images including 4 healthy classes, infected with powdery mildew, infected with black rot and infected with bacterial soft rot were prepared. Of these images, 70% were allocated for training data, 10% for validation data and 20% for test data.2D-LeNet convolutional neural network was used to analyze the images. This model was first unveiled in 1998 in an article called Gradient-Based Learning Applied to Document Recognition. The performance evaluation criteria of the classifiers are in the form of confusion matrix and receiver operating curve (ROC). The confusion matrix is the matrix that relates the real samples and the ones predicted by the classifier. In the following, some of these criteria will be examined

Sensitivity or readability:

How many percent of correct samples have been correctly diagnosed (Equation 2).

(2) Recall= $TP/(TP+FN)\times 100$

Accuracy:

Total percentage of correct answers of the system (Equation 3)

(3) Accuracy = $(TP+TN)/(TP+FN+FP+TN)\times 100$

Precision:

What percentage of correctly detected outputs are actually true (Equation 5).

(4) Precision=TP/(TP+FP)×100

F measure:

Harmonic weighted average of Recall and Precision (Equation 6)

(5) F measure=(2×Recall×Precision)/(Recall+Precision)

Where TP is the number of samples of each class that are correctly classified. TN is equal to the number of samples on the main diameter of the confusion matrix minus the number of samples that are correctly classified in the desired class. FN is defined as the sum of horizontal samples of the studied class minus the number of samples that are correctly classified in the desired class. And FP is the sum of the vertical samples of the examined class minus the number of samples that are correctly classified in the desired class.

Conclusion

The model was able to be trained with an accuracy of over 90% after the 15th epoch. After the 35th epoch, the validation and training graphs almost coincided. After the 35th epoch, the training was carried out with very high accuracy and minimum error and was validated. The results of the recall-accuracy displayed that 100% recall-accuracy index for each class shows the success of the classifier in identifying the mentioned 4 classes from each other. According to the receiver operating curve (ROC) for each class, it was stated that due to the fact that all the graphs are close to the line and the number one, it can be stated that all the classes were successfully identified and classified by the classifier.

At the final, the main results are as follows:

The model was able to be trained and validated with high accuracy of 99% and minimum error.

Healthy cauliflowers, cauliflowers infected with black rot and powdery mildew could be identified with 100% accuracy.

Cauliflower infected with bacterial soft rot could be identified with 99% accuracy

Keywords

Couliflower; CNN; Image processing; Disease.