VOLUMEN 48
XXXI Verano De la Ciencia
ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

Artificial intelligence as a tool to detect cells infected with baculovirus

Inteligencia artificial como una herramienta para detectar células infectadas con baculovirus

Michelle Dahian García Olavarria[1], Juana Isabel Macías Chávez[1], Jonatan Carmen Rangel Núñez[1,2], María de los Angeles Bivián Hernández[1,2], Alejandra Cecilia Franco Ramírez[1,2], María Cristina Del Rincón Castro[1,2*]

[1]Departamento de Alimentos
[2]Posgrado en Biociencias, División de Ciencias de la Vida, Campus Irapuato-Salamanca Campus, Universidad de Guanajuato. Km. 9.0 Carr. Irapuato-León, Ex Hacienda El Copal, C.P. 36500, Irapuato, Gto. México. cdelrincon@ugto.mx

Abstract

Baculoviruses are effective biological control agents of agricultural and forestry pests. The replication cycle of these viruses in insect cells consists of three stages: 3–6 hours post-infection (hpi) expresses immediate early genes, 12 hpi expresses late genes; and 24 hpi expresses very late genes. This process caused morphological changes in infected cells as: cell enlargement, growth of the nucleus, virogenic stroma formation; occlusion bodies (OB) formation in the cell nucleus, these OB accumulate and released upon cell lysis. These changes are not uniform across all cells simultaneously, making it difficult to detect infection in early stages. Using machine learning models, it is possible to determine baculovirus infection in cell lines. In this study, the Sf9 cell line was transfected with AcMNPV DNA, and the development of the infection was photographed every 24 hours for 5 days. The images were analyzed in Python using the libraries: mahotas, umap, and xgboost. The images were processed in OpenCV and NumPy. Dimensionality was adjusted using Principal Component Analysis (PCA) and the Uniform Collector Approximation/Projection (UCAP) algorithm. Machine learning (ML) models were trained, validated, and evaluated using the Random Forest, SVM, and XGBoost algorithms. The analysis and processing of the images showed that the homogeneity and confluence of infected cells remained consistent over time. By integrating the above information, the ML model differentiated between infected and uninfected cells with 88.99% accuracy from 48-120 hpi. Using ML an Artificial Intelligent model was generated that allows highly efficient prediction of the early presence of baculovirus-infected cells in insect cell lines.

Key words: Machin learning, Artificial Intelligence, Baculovirus, Transfection.

Introduction

In recent years, agricultural pests have become one of the main problems for food production. They cause losses of up to $220 billion and are responsible for 20–40% of global crop losses (Singh *et al*., 2024). The usual method for controlling these pests is the use of agrochemicals; however, these have been losing effectiveness due to the development of pest resistance (Banazeer *et al*., 2022; Singh *et al*., 2024). It makes necessary to implement the use of alternative control methods based on the use of natural enemies and pathogens of these pests (Chopra *et al*., 2025). Baculovirus are pathogens that affect agricultural pests and are recognized as biological control agents with the potential to keep pests under control, without affecting the ecological balance or harming non-target insects (Gelaye & Negash, 2023).

The mass production of baculoviruses is one of the main challenges for their implementation in the control of agricultural pests at field level. Mass production directly impacts production costs, increasing them. To address this, two production models have been proposed*: in vivo* production using whole insects and *in vitro* production using cell lines (Grzywacz *et al*., 2014; Reid *et al*., 2014). Each production method offers certain advantages and presents its own challenges.

Among the challenges faced by baculovirus production using insect cell lines for *in vitro* production is the observation that some baculovirus strains, particularly aggressive strains against fall armyworm (*Spodoptera frugiperda*), were unable to successfully complete their replication cycle and produce occlusion bodies (OB) in Sf9 cells by transfection method (Franco-Ramírez *et al*., 2025). Based on the above, in this work, the Sf9 cell line was transfected with DNA from the baculovirus model AcMNPV (*Autographa californica* multiple nucleopolyhedrovirus), and the development of the infection was documented using photographs. These imagens were used to train a machine learning algorithm and developed an Artificial Intelligence (AI) capable of detecting insect cells infected by a baculovirus. This will expedite through AI assistance, the process of detecting baculovirus infection in insect cell lines, accelerating research in this field and opening new frontiers for large-scale, cost-effective baculovirus production.

ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

Materials and methods

Viral strain

The baculovirus strain used was AcMNPV described by Jiménez-Hernández *et al*., 2022. The virus was amplified using second instar *S. frugiperda* larvae using the diet surface contamination technique in Petri dishes (90x15 mm) with artificial diet, in which 1x107 OB was dissolved in 1 mL of sterile distilled water (SDW) and placed per dish. 10 S. frugiperda larvae were placed per dish and incubated for 7 days under controlled conditions, after which they were collected and stored at -20 °C.

Purification of viral strain

Infected larvae were processed according to Rangel-Nuñez *et al*., 2026 using a mortar and a solution of 0.2% SDS. The OB was suspended in SDW and purified in sucrose gradients 66-40% weight/weight in an Ultracentrifuge Beckman Coulter (XPN-100) at 24,000 rpm, 1.5 hour and 4 °C. The OB band was collected and dissolved in SDW, the solution was centrifuged at 14,000 rpm, 10 minutes. The supernatant was discarded and the pellet was washed with 10 mL of SDW. The process was repeated three times. Finally, the OB was dissolved in 5 mL of SDW and quantified according to Zanella-Saenz *et al*., 2022. The OB was stored at -20°C.

DNA extraction

DNA was extracted from AcMNPV as reported by Franco-Ramírez *et al*., 2026 using 1x1011 OB treated with alcali (0.1 M Na2CO3, 0.1 M NaCl, pH 11), proteinase k digestion and Fenol:Chloroform:Isoamyl alcohol treatment. The DNA was dissolved in TE (10mM Tris, 1mM EDTA pH 8.0) and quantified using NanoDrop Lite (Thermo Scientific).

Transfection of Sf9 cell line

Once the DNA was obtained the cellular line was transferred. 2x106 Sf9 lived cells were seed in 6-well multiplate (Biolite) in Sf900 III media without fetal bovine serum. Then transfection solution was prepared mixing 950 µL of HBS solution (20 mM HEPES, 1 mM NaHPO4, 5 mM KCl, 140 mM NaCl, and 10 mM glucose, pH7.05) with 20 µL of ExpiFectamine (Gibco) in an Eppendorf tube, the solution was mixed manual for 1 minute. Then 1 µg of AcMNPV DNA was added. The solution was to stand for 3 minutes at room temperature. After this time the Sf900 III media was removed from cells, and the transfection solution was added drop by drop. This solution was left in contact with cells for 1 hour, after this time the solution was removed, and fresh Sf900 III medium was added. The cells were stored at 28 °C and photographed (nine photos per treatment) (Iphone 15 pro max RAW48 mega pixels resolution) in inverted microscopy (Primo Vert Zeizz) every 24 hours until OB was observed.

Image obtention

Once the images of the Sf9 cells transfected with AcMNPV DNA were obtained, the images were processed. A pipeline was developed in Google Colab that allowed for the comparative analysis of images of Sf9 cells transfected with AcMNPV baculovirus and uninfected Sf9 cells. Subsequently, the image processing was divided into five stages: Installing libraries and loading images, image preprocessing, feature extraction and dataset construction, exploratory data analysis and binary classification.

Library Installation and Image Loading

Three Python libraries were used: mahotas, umap, and xgboost. The first, mahotas, was used for image processing and computer vision; its interface is divided into seven categories. In this research, Haralick texture features were used to obtain statistical properties such as entropy, homogeneity, contrast, and energy from

the grayscale-transformed photographs (Coelho, 2013). Subsequently, the diversity of cell morphology associated with baculovirus infection was examined through exploratory analysis. Scatter plots were generated to determine the clustering of infected and uninfected cells during the study period using umap (Becht *et al*., 2019). Finally, the xgboost library created the predictive model of the parameters, under it the program determined if Sf9 cells were infected and the progression of the infection at different times.

Image preprocessing

Using the OpenCV and NumPy libraries, the imagens of infected and uninfected Sf9 cells were converted to grayscale, performed a lighting correction, and the contrast of each of the photos was increased (CLAHE) (Rajab Asaad *et al*., 2023).

Feature extraction and dataset construction

For each image, three types of features were evaluated: morphological, texture and intensity, each divided into different categories (Table 1).

Table 1. *Characteristics evaluated in Sf9 cells transfected with baculovirus DNA and uninfected Sf9 cells.*

Characteristic	Name	Abbreviation
Morphology	Percentage of area occupied	morf_area_ocupada_pct
	Number of objects	morf_num_objetos
	Average size of objects	morf_tamano_promedio
Texture	Entropy of Shannon Global	text_entropia
	Descriptors of Haralick	
	Contrast of Haralick	haralick_contraste
	Homogeneity of Haralick	haralick_homogeneidad
	Energy of Haralick	haralick_energia
	Correlation of Haralick	haralick_correlación
	Local binary patterns LBP	lbp_bin_0 al lbp_bin_9
Intensity	Average intensity	int_media
	Standard deviation of intensity	int_std
	Intensity histogram	int_hist_bin_0 al int_hist_bin_7

Exploratory Data Analysis

To facilitate data understanding, visualization, and improve the performance of the models used, dimensionality reduction (principal components) was performed using Principal Component Analysis (PCA) and the Uniform Collector Approximation and Projection (UMAP) algorithm. Correlations and redundancies between the studied variables (Table 1) were determined using Pearson correlation coefficients and visualized through heat maps. Cell area and entropy over the five days of the study were visualized using box plots (Healy & McInnes, 2024).

Binary Classification

In this stage, machine learning models were trained, validated, and evaluated using the Random Forest, SVM, and XGBoost algorithms. These algorithms help train forests of trees (100) according to the analyzed variables (Table 1), selected randomly. Each tree classifies cells as infected or uninfected. If the first forest of trees

misclassifies any image, the second attempts to correct the errors; the third corrects the second, and so forth (Abbas & Dijkstra, 2020). 70% of the data from infected and uninfected Sf9 cells were used to train the algorithms, and the remaining 30% was used as test set. From this code cell, a classification report was obtained, and evaluated three metrics: accuracy, precision, and recall. The first parameter refers to the percentage of correctly classified images. The accuracy refers of all the photos that the model "said were infected" how many of they really were infected. And the sensitivity (recall) refers to all the photos that were dictamined as infected how many were detected by the model.

Results and Discussion

Transfection of the Sf9 Cell Line

To induced baculovirus infection symptoms in insect cells, the Sf9 cell line was transfected with DNA of the standard baculovirus AcMNPV. To achieve this, the DNA of AcMNPV was extracted from OB (Fig. 1A, Lane 1). Using 1,000 ng of this DNA, the Sf9 cell line was transfected. The transfection was observed in Fig. 1B. At 24 hours, no differences were observed between the transfected and uninfected cells (Fig. 1B). Differences between the transfected and uninfected cells were observable from 48 hours onward. At that time, the transfected cells showed cells with atrophied nuclei and the presence of virogenic stroma (blue arrows) (Fig. 1B). These signs of infection were completely visible in the transfected cells at 72 hours (Fig. 1B). At that time, the first cells with OB were observed (red arrows) (Fig. 1B). The presence of cells with OB mixed with cells with virogenic stroma was predominant between 96 and 120 hours (Fig. 1B). Between 72 to 120 hours cell confluence increased significantly in the uninfected cells; at 120 hours cell confluence difficulted observe cell morphology with the naked eye. This confluence was not observed in the transfected cells at any time point.

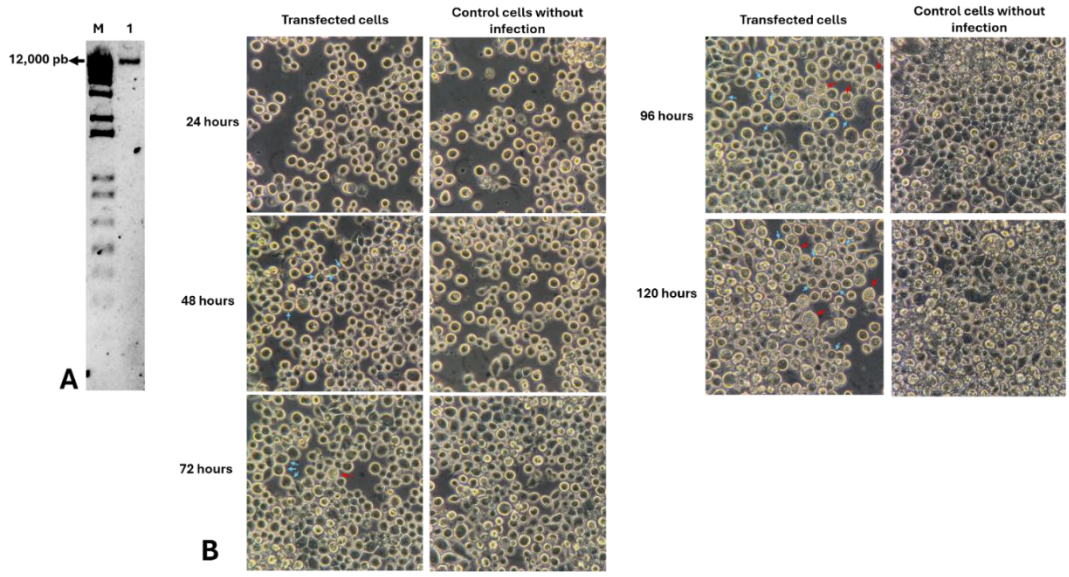


Figure 1. DNA extraction and Sf9 cell line transfected at different time post infection. A) DNA extraction. Lane 1 DNA of AcMNPV. B) Transfection of Sf9 cellular line. Left, cells transfected. Right, control cells without infection. Blue arrows indicated cells with symptoms of infection and without OB. Red arrows indicate cells infected with the virus and OB in the nucleus.

Machine learning training

Correlation of variables

Once the images of the transfected and uninfected cells were obtained, a correlation matrix was created. Figure 2 shows the correlation matrix between the study variables (morphology, texture, and intensity) of the transfected Sf9 cells and uninfected cells. In the upper left of Fig. 2 (intense red), it was observed that the standard deviation of intensity (int_std), global entropy (text_entropia, haralick_contrast, and

haralick_entropia), intensity (int_his_bin_5 to 7), and texture (lbp_bin_4 and lbp_bin_6) were grouped. These results indicated the positive relationship between the variables to each other. At the cellular level, the analysis of the variables reinforces the visual observations of the infection progresses across the time, which caused a visual disorder or entropy. Entropy is caused by changes in the shape, size, and confluence of infected cells (Üstün-Aytekin *et al*., 2014). The development of nuclear hypertrophy, virogenic stroma, and, in more advanced stages, the presence of occlusion bodies in cell nuclei and cell lysis were observed in transfected cells, similar to what was observed by Zanella *et al*., 2022. The development of the infection in cells and the morphological changes altered the pixels in the imagens and in advanced stages of infection, the presence of light-refracting occlusion bodies affected the brightness captured in the images; this was considered positive because it facilitated the detection of changes in the transfected Sf9 cell line, in contrast to the uninfected cells.

In Fig. 2, the characteristics recorded in the blue block were: haralick_homogeneity, int_his_bin_3 and 4, and textures such as lbp_bin_7 and 8. These showed that as the baculovirus infection progressed in the Sf9 cells, the homogeneity was maintained in infected cells and decreased in healthy cells. In contrast to observed in the red block, visual disorder increased in uninfected cells.

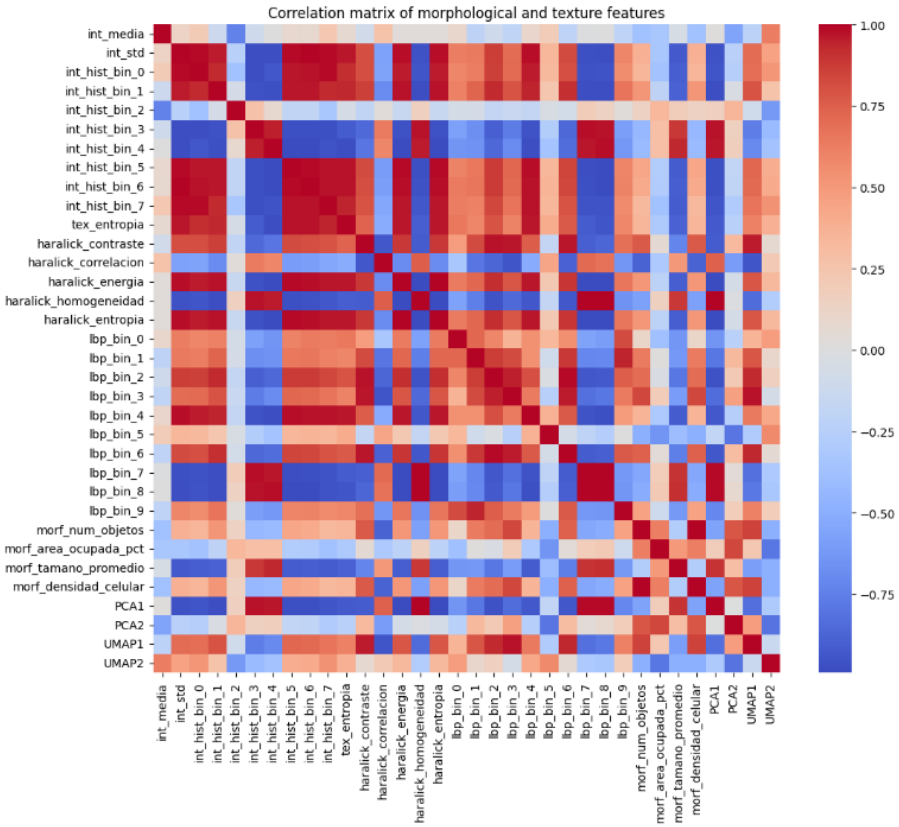


Figure 2. Pearson correlation matrix heatmap.

Figure 3 shows variations in the analyzed characteristics at different times between transfected and uninfected cells. At 24 hours (purple), a clustering was observed between both conditions; at this point, the algorithm did not detect significant differences between the transfected and healthy cell lines (Fig. 3). Between 48 and 120 hours (blue, dark green, light green, and yellow), the dispersion observed between transfected and healthy cells was more evident, indicating the beginning of detectable changes in the transfected Sf9 cells. These results were consistent with those reported by Jiménez-Hernandez *et al*., 2022, they observed the presence of occlusion bodies starting at 36 hours in Sf9 cells transfected with AcMNPV baculovirus DNA. The dispersion of points for both conditions over the five days demonstrated that nuclear hypertrophy, granularity, and cell lysis were changes that can be efficiently measured using this technique.

Two characteristics of the Sf9 cell culture were evaluated: confluence and entropy. Figure 4A shows that the uninfected cells (purple) exhibited a decrease in growth, but at 120 hours growth dramatically increased, indicating that the cells continued dividing until they achieved greater confluence. In contrast, transfected cells (green) stopped dividing after 24 hours; at 120 hours, a slight increase in cell division was observed; however, the cell area occupied was considerably lower. The data observed in Figure 4A demonstrates that baculovirus infection inhibited cell proliferation, reaching its maximum effect at 96 hours. This coincided with the late/very late phase where lysis begins, affecting the cells' confluence.

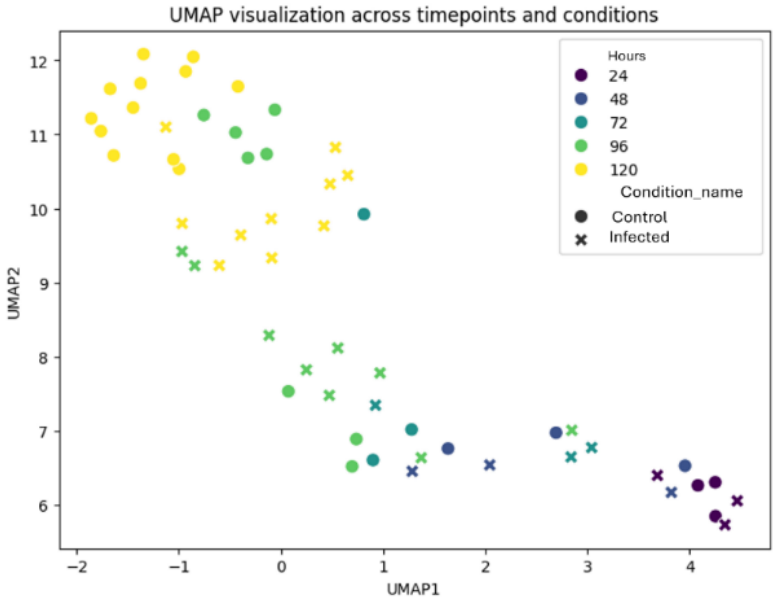


Figure 3. Evaluation plots by time and condition.

Figure 4B compares the level of disorder in both conditions: transfected and uninfected cells. The evaluation in this image about roughness or heterogeneity detectable in the grayscale of the images. It was observed that, over time, both conditions showed an increase in entropy; however, the most evident change occurred at 72 hours. In uninfected cells (yellow), entropy still increased, while in infected cells (blue) entropy decreased. These results show that, between transfected and uninfected cells, the entropy increased and persists over time due to the natural aging of cells, as well as the lack of synchronized transfection of cells. The confluence increased over time, affecting the visual characteristics of the images. At 72 hours the difference in entropy between the control and transfected cells was evident (Fig 4B), this was because the cytoplasm of the Sf9 cells suffered a readjustment of the cytoskeleton of the transfected cells and the massive production of viral proteins.

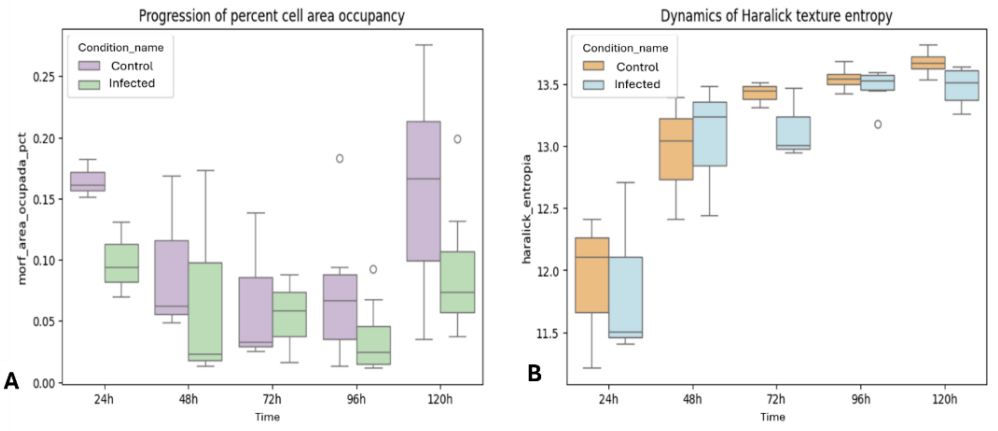


Figure 4. A) Comparison of cell confluence between infected and uninfected cells. B) Comparison of the level of disorder between infected and uninfected cells.

Binomial Classification

The most robust model for classifying images of Sf9 cells transfected with AcMNPV DNA was Random Forest, with an accuracy of 88.99% (Table 2). This result is important because indicated that the model was accurate and presented few false negatives and few false positives. This was corroborated by testing the model. In Fig. 5 the confusion matrix was observed, this indicated the total number of images that were correctly classified. In Fig. 5, it was observed that eight uninfected cell images were correctly classified, and only one was classified as transfected cells. Similarly, eight of the images were classified as transfected cells, and only one was classified as uninfected cells.

Table 2. *Values for the classification by the Random Forest method.*

	Precision	Sensibility	F1-score	Support
Uninfected	0.89	0.89	0.89	9
Transfected	0.89	0.89	0.89	9
Precision			0.89	18

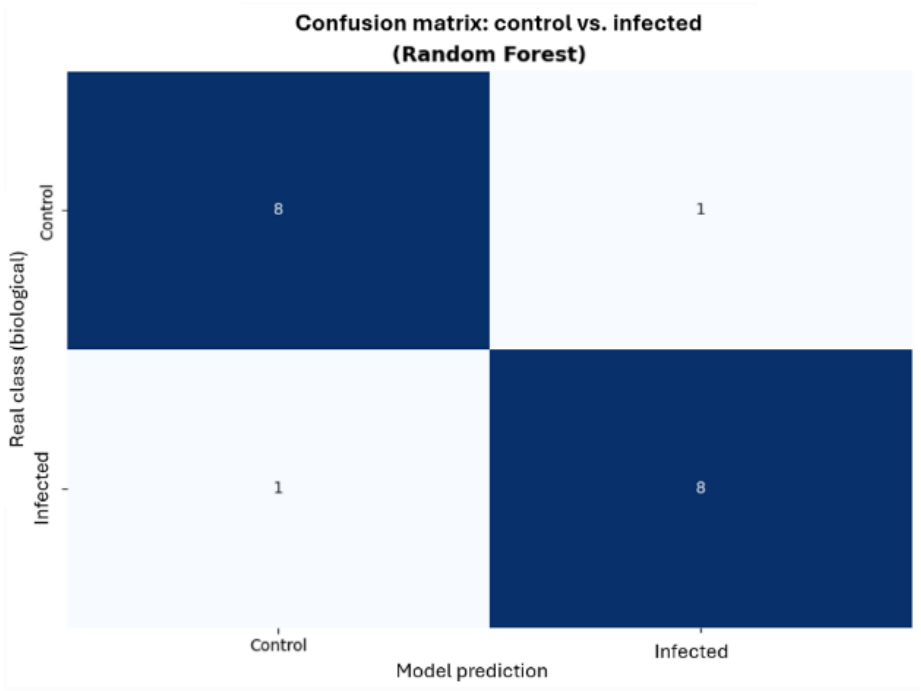


Figure 5. Confusion matrix for Sf9 cells infected with AcMNPV baculovirus and uninfected cells.

Conclusions

In this study, an experiment was designed in which the Sf9 cells derived from S. frugiperda were transfected with the DNA of the baculovirus AcMNPV. The development of changes in the structure of the cells during the evolution of the transfection in permissive cells was documented in imagens. Using the imagens, an IA model was trained using machine learning to differentiate between transfected cells with a baculovirus and uninfected cells. The training of IA model was successful, and the model was able to distinguish between imagens of cells transfected with baculovirus and uninfected cells with high efficiency in 89% of the cases analyzed. The result of this work contributes to the development of new strategies based on the use of AI technologies to detect viral infections in insect cells. With this AI capable of identified cells infected with a virus the expansion in basic science research of the cellular phenomena associated with baculovirus infection will grow up, generating new knowledge assisted by technological tools.

References

Abbas, S. S., & Dijkstra, T. M. H. (2020). Detection and stage classification of Plasmodium falciparum from images of Giemsa stained thin blood films using random forest classifiers. *Diagnostic Pathology*, 15(1). https://doi.org/10.1186/s13000-020-01040-9

Banazeer, A., Afzal, M. B. S., Hassan, S., Ijaz, M., Shad, S. A., & Serrão, J. E. (2022). Status of insecticide resistance in Plutella xylostella (Linnaeus) (Lepidoptera: Plutellidae) from 1997 to 2019: cross-resistance, genetics, biological costs, underlying mechanisms, and implications for management. *Phytoparasitica*, 50(2). https://doi.org/10.1007/s12600-021-00959-z

Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., & Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, 37(1), 38–47. https://doi.org/10.1038/nbt.4314

Chopra, M., Choudhary, M., Singh, M., Jat, H. S., Tureevna, Z. L., Sadykova, N., & Kumar, V. (2025). Biocontrol strategies in agricultural green development and soil health management to achieve sustainable growth in agriculture. *Discover Applied Sciences*, 7(10). https://doi.org/10.1007/s42452-025-07371-w

Coelho, L. P. (2013). Mahotas: Open-source software for scriptable computer vision. *Journal of Open Research Software*, 1(1), e3. https://doi.org/10.5334/jors.ac

Franco-Ramírez, A. C., Rangel-Núñez, J. C., Bivián-Hernández, M. A., & Rincón-Castro, M. C. (2025). Transfección de línea celular sf9 de Spodoptera frugiperda con cepas nativas de baculovirus. *Revista Ciencia e Innovación Agroalimentaria de La Universidad de Guanajuato*, 6, 101–111. https://doi.org/10.15174/cia.v6i6.103

Gelaye, Y., & Negash, B. (2023). The role of baculoviruses in controlling insect pests: A review. *Cogent Food & Agriculture*, 9(1). https://doi.org/10.1080/23311932.2023.2254139

Grzywacz, D., Moore, D., & Rabindra, R. J. (2014). Mass Production of Entomopathogens in Less Industrialized Countries. *Mass Production of Beneficial Organisms: Invertebrates and Entomopathogens*, 519–561. https://doi.org/10.1016/B978-0-12-391453-8.00015-7

Healy, J., & McInnes, L. (2024). Uniform manifold approximation and projection. *Nature Reviews Methods Primers*, 4(1). https://doi.org/10.1038/s43586-024-00363-x

Jiménez-Hernández, S. Y., Rangel-Núñez, J. C., Ibarra, J. E., & Del Rincón-Castro, M. C. (2022). Biological, morphological, and molecular characterization of the baculovirus PlxyMNPV_LBIV-11, and its virulence towards Plutella xylostella, Trichoplusia ni, and Spodoptera frugiperda larvae. *Archives of Microbiology*, 204(10). https://doi.org/10.1007/s00203-022-03222-z

Rajab Asaad, R., Ismael Ali, R., Arif Ali, Z., & Ahmad Shaaban, A. (2023). Image Processing with Python Libraries. Academic *Journal of Nawroz University*, 12(2), 410–416. https://doi.org/10.25007/ajnu.v12n2a1754

Rangel-Nuñez, J. C., Bivian-Hernández, Ma. de los A., & Del Rincón-Castro, Ma. C. (2026). Caracterización de las proteínas de los cuerpos de oclusión de baculovirus con actividad hacia el gusano cogollero del maíz Spodoptera frugiperda. *Mexican Journal of Biotechnology*, 11(1), 21–48. https://doi.org/10.29267/mxjb.2026.11.1.21-48

Reid, S., Chan, L., & van Oers, M. M. (2014). Production of Entomopathogenic Viruses. *Mass Production of Beneficial Organisms: Invertebrates and Entomopathogens*, 437–482. https://doi.org/10.1016/B978-0-12-391453-8.00013-3

Singh, A., Shraogi, N., Verma, R., Saji, J., Kumar Kar, A., Tehlan, S., Ghosh, D., & Patnaik, S. (2024). Challenges in current pest management practices: Navigating problems and a way forward by integrating controlled release system approach. *Chemical Engineering Journal*, 498, 154989. https://doi.org/10.1016/J.CEJ.2024.154989

Üstün-Aytekin, Ö., Gürhan, I. D., Ohura, K., Imai, T., & Öngen, G. (2014). Monitoring of the effects of transfection with baculovirus on Sf9 cell line and expression of human dipeptidyl peptidase IV. *Cytotechnology*, 66(1), 159–168. https://doi.org/10.1007/s10616-013-9549-3

Zanella, S., Herniou, E., Ibarra, R., Huerta, A., & Del Rincón, C. (2022). Virulence and genetic characterization of six baculovirus strains isolated from different populations of Spodoptera frugiperda (Lepidoptera: Noctuidae). *Archives of Microbiology*, 204(1). https://doi.org/10.1007/s00203-021-02722-8