Prediction of the Change Rate of Tumor, Healthy Host, and Effector Immune Cells in a Three-Dimensional Cancer Model using Extended Kalman Filter

Nina Fitriyati, Salma Abidah Faizah, and Taufik Edy Sutanto
Prediction of the Change Rate of Tumor, Healthy Host, and Effector Immune Cells in a Three-Dimensional Cancer Model using Extended Kalman Filter

Nina Fitriyati1,*, Salma Abidah Faizah2, and Taufik Edy Sutanto3

1. Introduction

Cancer is one of the most deadly diseases despite advances in science and medicine. This disease escalates due to changes in the formation of DNA (mutations) of abnormal cells. Uncontrolled cell proliferation leads to tumors, and many mutated cells occur daily in the human body. These mutated cells form cancer cells. Cancer can also occur when the immune system or other defense mechanisms fail to protect the human body from these cells [1]. Cancer cells differ from normal cells in size, shape, number, differentiation, function, and ability to travel to distant tissues and organ systems [1]. The immune system recognizes cancer cells and tumors based on their antigens [2]. This complex biological phenomenon can be accurately described using mathematical modeling.

In current literature, effective mathematical modeling significantly contributes to understanding and analyzing diseases, including tumor growth. Therefore, it is crucial to study computational models of how the immune system interacts with other factors. One mathematical model that captures the interaction between tumor growth and the immune system is the three-dimensional cancer model studied by [3]. This model describes the dynamics of three interacting cell populations: tumor, healthy host, and effector immune cells. Itik and Banks [3] explain the biological relevance and analyze the model’s equilibrium.

The rate of change of tumor cells, healthy host cells, and effector immune cells is crucial in understanding and managing cancer dynamics. Tumor growth is significantly influenced by the type and state of the tissue, as well as environmental factors, which can explain the varying prevalence of cancers in different tissues [4]. The interaction between tumor cells and effector immune cells is complex, with models showing that increased activation of effector cells can lead to chaotic dynamics and potentially facilitate cancer clearance [5]. The immune system’s role is further highlighted by the effectiveness of therapies that target tumor stroma, where effector cells like CD8+ CTLs can inhibit tumor growth by destroying immunosuppressive stromal cells [6].

Mathematical models, such as the NTIUNHDM, demonstrate that a weak immune response can lead to the coexistence of normal and abnormal cells, emphasizing the importance of early immune system activation to reduce cancer risk [7].

Additionally, the stochastic nature of tumor-immune interactions suggests that different noise intensities in the microenvironment can lead to varying outcomes, including tumor extinction or persistence [8]. The role of effector cells, including T lym-
phocytes and dendritic cells, is critical in mediating anti-tumor responses, with their activity being influenced by factors such as diet and immune system strength [9, 10]. Furthermore, the non-ischemic effects of treatments like flavone acetic acid (FAA) are mediated by immune effector cells, indicating that immune responses can significantly impact tumor cell dynamics independent of blood flow [11]. Overall, the interplay between tumor cells, healthy host cells, and effector immune cells is a dynamic process influenced by various biological and environmental factors, and understanding these interactions is critical to developing effective cancer therapies [12, 13].


The Itik-Banks model is a nonlinear and chaotic dynamical system [21]. One suitable method for estimating the state is the Extended Kalman Filter (EKF) [22]. The Extended Kalman Filter (EKF) offers several advantages over other estimation methods, such as the Unscented Kalman Filter (UKF), Particle Filter (PF), and Ensemble Kalman Filter (EnKF). One of the primary benefits of the EKF is its ability to handle nonlinear systems by linearizing them around the current estimate using the Taylor series expansion, which makes it computationally efficient and suitable for real-time applications [23]. While potentially introducing linearization errors, this linearization approach is often sufficient for many practical applications, such as vehicle state estimation and structural health monitoring, which have been validated through experiments [24, 25]. In contrast, the UKF, while avoiding linearization by using sigma points to capture the mean and covariance accurately, requires careful tuning of parameters, which can be cumbersome [26]. The Particle Filter (PF), including variants like the EnKS-MDE or, provides a more flexible approach by approximating the state probability density function using multiple conditional state pdfs. Still, it can suffer from particle degeneracy and is computationally intensive [26, 27].

Particle Filters (PF), including the Ensemble Kalman Filter (EnKF) and the Ensemble Kalman Smoother Multiple Distribution Estimation Filter (EnKSmDE), offer superior accuracy by representing the state probability density function with multiple samples or particles. Still, they require many samples, leading to high computational costs and potential degeneracy issues [28, 29]. The EKF, despite its limitations in handling strong nonlinearities, remains advantageous in scenarios where computational efficiency and simplicity are paramount, such as in satellite attitude estimation and real-time control systems. Moreover, the EKF’s ability to balance computational efficiency and estimation accuracy makes it a preferred choice in many practical applications, including industrial systems and robotics, where real-time performance is critical. Thus, while the UKF, PF, and EnKF offer higher accuracy and robustness in certain conditions, the EKF’s lower computational requirements and ease of implementation provide significant advantages in resource-constrained environments [30, 31].

Based on several advantages of EKF, in this study, we seek to develop and implement the Extended Kalman Filter (EKF) method to forecast the rate of change in tumor cells, healthy host cells, and effector immune cells within the Itik-Banks model. To date, no research has employed the EKF method for this specific application, positioning this study as an innovative and novel approach in the modeling of cancer dynamics. The application of EKF is anticipated to yield more precise and real-time estimations of the intricate cellular dynamics described by the Itik-Banks model, thereby offering significant contributions to the comprehension and treatment of cancer. A state space is formed by linearizing the Itik-Banks model using the first-order Taylor series. Numerical simulations are performed using data from 2000 rates of change in tumor, healthy host, and effector immune cells sourced from William Gilpin’s GitHub repository. Predictions of the three cell types are made when and when data is available. This paper is organized as follows: section 2 describes the three-dimensional cancer model and the EKF algorithm. Section 3 gives the linearization of the Itik-Banks model, forming state space, and simulation of the prediction tumor, healthy host, and effector immune cells for training and testing data, and section 4 concludes.

2. Methods

2.1. Three-Dimensional Cancer Model

This article uses the three-dimensional cancer dynamical model, i.e., the Itik and Banks model [3]. This model describes the dynamics of three interacting cell populations: tumor, healthy host, and effector immune cells, namely cytotoxic T lymphocytes CD8, CTL. Effector cells are active cells with relatively short lifespans from the body’s immune system that defend the body in immune responses. Similar to previous cancer models [32–37], this model depicts the competition dynamics among the three types of interacting cells in a well-mixed system (e.g., liquid cancers like leukemia or multiple lymphomas). Among several biologically significant assumptions (see [3]), this model assumes that the anti-tumor effect of the immune system response is carried out by cytotoxic T cells, i.e., those mediated by adaptive T cell-based groups. Alpha-beta T cells are activated after their T cell Receptor (TCR) recognizes tumor-specific antigens on the cell surface in small peptides presented in the context of major histocompatibility complex (MHC) molecules. CD8 T cells are responsible for direct cell-mediated cytotoxicity after activation by antigen-presenting cells (APCs) and are considered central players in anti-tumor immune responses. To achieve full activation, signals from the TCR must be reinforced with messages sent by costimulatory molecules such as CD28, which are also present on the surface of T cells. Failure to engage costimulatory proteins, activation of coinhibitory receptors such as CTLA-4 or PD-1, or the presence of regulatory CD4 (TReg) T cells can lead to failure of T cell activation or decreased regulation of immune responses.
Uncovering these inhibitory mechanisms can lead to the reactivation of anti-tumor immune responses and supraphysiological levels of T cell activation, which is helpful in clinical management.

The Itik-Banks tumor model consists of three types of cell populations: $T(t)$ is the count of the number of tumor cells in time $t$, $H(t)$ is the number of healthy host cells in time $t$, and $E(t)$ is the number of effector immune cells in time $t$. The interaction between these three cells follows the nonlinear differential equation system [3]:

$$\frac{dT(t)}{dt} = r_1 T(t) \left( 1 - \frac{T(t)}{k_1} \right) - a_{12} T(t) H(t)$$

$$\frac{dH(t)}{dt} = r_2 H(t) \left( 1 - \frac{H(t)}{k_2} \right) - a_{21} H(t) T(t)$$

$$\frac{dE(t)}{dt} = r_3 T(t) E(t) - a_{31} E(t) E(t) - d_3 E(t)$$

subject to initial condition $(T(0), H(0), E(0)) = (T_0, H_0, E_0)$ $\geq 0$ and $a_{12}, a_{21}, a_{31}, a_{13}, a_{12}, a_{21}$, and $k_3$ are constants. For simplicity of analysis, we non-dimensionalize equations (1)-(3) by employing the definitions:

$$T(t) = \frac{T(t)}{T_0}, \quad H(t) = \frac{H(t)}{H_0}, \quad E(t) = \frac{E(t)}{E_0}, \quad t = r_1 \tau.$$

A new parameters following eq. (4) are

$$a_{31} = \frac{a_{13} k_1}{r_1}, \quad a_{13} = \frac{a_{12} k_3}{r_1}, \quad a_{12} = \frac{a_{12} k_2}{r_1},$$

$$a_{31} = \frac{r_1}{k_1}, \quad r_2 = \frac{r_2}{r_1}, \quad r_3 = \frac{r_3}{r_1},$$

$$k_3 = \frac{k_3}{k_1}, \quad d_3 = \frac{d_3}{r_1}.$$  

Using eq. (4) and eq. (5), the dimensionless system of differential equations for the Itik-Banks tumor is as follows [3]:

$$\dot{T}(\tau) = \frac{dT(t)}{dt} = T(t)(1 - T(t)) - a_{12} T(t) H(t)$$

$$\dot{H}(\tau) = \frac{dH(t)}{dt} = r_2 H(t)(1 - H(t)) - a_{21} T(t) H(t),$$

$$\dot{E}(\tau) = \frac{dE(t)}{dt} = r_3 T(t) E(t) - a_{31} E(t) E(t) - d_3 E(t).$$

In equations (6)-(8), the parameter $a_{12}$ is the rate of tumor cell inactivation by healthy cells; $a_{13}$ is the rate of tumor cell inactivation by effector cells; $r_2$ is the intrinsic growth rate of healthy tissue cells; $a_{12}$ is the rate of inactivation of healthy cells by tumor cells; $r_3$ corresponds to the activation rate of effector cells due to the recognition of tumor cell antigens; $a_{13}$ is the rate of effector cell inactivation by tumor cells, and $d_3$ is the density-dependent effector cell death rate [3].

The last term in equation (6) indicates the inactivation rate (or elimination) of tumor cells by the action of effector immune cells assumed to be proportional to the number of effector immune cells and does not consider saturation. The mechanism of tumor cell elimination is provided through the release of cytotoxic granules by effector cells that damage or destroy tumor cells. Effector cells can proliferate clonally after antigen recognition, so the model assumes these cells can be in excess if needed. The activation of effector immune cells due to antigen recognition used in the first term of equation (8) can be seen as a functional response Holling-II, commonly used to model predator feeding saturation in dynamic ecological systems. Holling Type II functional response holds biological significance in stimulating the immune response, as evidenced by various research papers. It is utilized in models studying autoimmune disorders like Guillain-Barre syndrome [38], predator-prey dynamics with diseases in the predator population [39], and the impact of deforestation on wildlife species [40]. The Holling Type II functional response is particularly effective in describing the predation of susceptible and infected predators and the interaction between wildlife and forest resources.

Using data from William Gilpin’s GitHub repository, i.e., the Itik-Banks tumor. Gilpin provides 131 ordinary differential equation (ODE) models for the dynamics of disease spread and resource utilization in ecological systems, making it a valuable tool for understanding and controlling epidemics and ecosystem stability. For the above system, it is assumed that there is a slowing down of the activation rate as the number of tumor cells increases because the activation of effector immune cells is limited by the need for these cells to recognize tumor antigens in the context of Antigen Presenting Cells (APCs). Cell interaction processes and receptor recognition are required between APCs and tumor cells before activation. Thus, increasing the number of tumor cells does not always imply an increase in effector cell activation [20].

The Itik and Banks model is a chaotic dynamic system [21]. A chaotic dynamical system is a nonlinear dynamic system with unique characteristics. The behavior of this system is sensitive to initial conditions, and it is quite challenging to predict. Dynamic systems are often used to model complex systems’ behavior, which is highly uncertain. We can use the maximum Lyapunov exponent to determine whether a system is chaotic and sample entropy to observe the system complexity [41].

Maksimum Lyapunov exponent, $\lambda$, is calculated using [42]:

$$\lambda = \left( \frac{1}{n^2} \right) \sum_{i} \ln \left( \frac{D(i,j)}{\overline{D}} \right),$$

where $\overline{D} = \frac{1}{n^2} \sum_{i \neq j} ||x(i) - x(j)||$ is the average distance of all pairs of points, $D(i,j)$ is the distance between point $x(i)$ and $x(j)$, and $n$ is the time distance.

Assume we have a time-series data set of length $N = \{x_1, x_2, ..., x_N\}$ with a constant time interval $\tau$ define a template vector of length $m$, such that $X_m(i) = \{x_i, x_{i+1}, x_{i+2}, ..., x_{i+m-1}\}$ and the distance function $d_i(X_m(i), X_m(j))$ for $i \neq j$ is to be the Chebyshev distance. The sample entropy is calculated using the formula eq. (10) [43], and the complexity classification based on SampEn values is shown in Table 1 [44].

$$SampEn = -\ln \left( \frac{d(m+1,r)}{d(m,r)} \right),$$

where $r$ is a positive number that sets the tolerance level for matching.

2.2 Data

We use data from William Gilpin’s GitHub repository, i.e., the Itik-Banks tumor. Gilpin provides 131 ordinary differen-
The equations (6)-(8) display a nonlinear relationship between the tumors, healthy hosts, and effector immune cells. This system is chaotic [21]; hence, the Extended Kalman Filter (EKF) is an appropriate model to estimate state [22]. The EKF adapts the linear Kalman filter used for systems with nonlinear measurements and state equations [45]. In the EKF algorithm, nonlinear equations are transformed into linear equations through linearization, which involves using a first-order Taylor series. This leads to a linear system that can be used with the Kalman filter algorithm [46]. The state-space representation in EKF is expressed as follows:

\[ x_t = f(x_{t-1}) + w_t, \]
\[ z_t = h(x_t) + v_t, \]

Equation (11) is called the state equation, and equation (12) is called the measurement equation. The nonlinear functions \( f \) and \( h \) define the state transition function and the relationship between state and measurement. Vector \( x_t \) and \( z_t \) are the state and measurement vector, respectively, and \( w_t \) and \( v_t \) are the state error and measurement error, respectively. State error refers to the difference between the true state of the system and the estimated state, and measurement error refers to the difference between the actual measured output of the system and the true output. State error helps in evaluating the accuracy of the state estimation process [47]. Measurement error is important for system identification and parameter estimation, significantly impacting the accuracy and reliability of the results [48]. Assume that \( E[w_t] = 0, \text{cov}(w_t) = Q_t, E[v_t] = 0 \) and \( \text{cov}(v_t) = R_t \).

The nonlinear functions \( f \) and \( h \) are linearized using a first-order Taylor series as follows:

\[ x_t = \Phi_{t-1} x_{t-1} + w_t, \]
\[ z_t = H x_t + v_t, \]
\[ \Phi_{t-1} = \frac{\partial f(x_{t-1})}{\partial x_j}, \]
\[ H_t = \frac{\partial h_i(x_t)}{\partial x_j}, \]

where \( f \) and \( h \) are element \( i^{th} \) from \( f \) and \( h \), respectively, \( \Phi \) and \( H \) are state transition and measurement matrices, respectively.

Like the Kalman Filter, state estimation in the EKF algorithm has two steps: forecast (prior) and update (posterior). The forecast step is to predict the current state given the previous state and the time that has passed since. The update step combines the predicted state with an incoming measurement that involves the Kalman gain, \( K_t \). State and covariance in the forecast step will be symbolized by \( x_t \) and \( P_t \), and state and covariance in the update step will be symbolized by \( x_t \) and \( P_t \). The pseudocode from the EKF algorithm follows [46]:

**Algorithm 1 Extended Kalman Filter**

Input: \( x_{t-1}, P_{t-1}, z_t \)

Output: \( x_t, P_t \)

1. **Forecast/Prior Step**
   \[ x_t^- = \Phi_{t-1} x_{t-1} \]
   \[ P_t^- = \Phi_{t-1} P_{t-1} \Phi_f + Q_t \]

2. **Update/Posterior Step**
   \[ K_t = P_t^- H f (H P_t^- H^T + R_t)^{-1} \]
   \[ x_t \sim x_t^- + K_t (z_t - H x_t^-) \]
   \[ P_t \sim (I - K_t H_t) P_t^- \]
   return \( x_t, P_t \).

The Extended Kalman Filter (EKF) is a powerful tool that has been successfully implemented in various fields, including radar-based object tracking [46, 49, 50], traffic [51–53], financial market [54], stock price [55, 56], and robotic problem [57, 58]. Its versatility and efficiency make it an indispensable asset for anyone needing accurate and reliable data analysis.

4. Results and Discussion

The Itik and Banks model describes the dynamics of three interacting cell populations: tumors, healthy hosts, and effector immune cells. Figure 1 shows the training data from the change rate of these three cells. Figure 2 shows the correlation coefficients between these three cells. From Figure 1, we see that tumor and immune cells’ change rate patterns are similar. Still, these two cells have almost no linear relationship because the correlation coefficient value for both cells is very small. While tumor cells have a fairly strong negative relationship with healthy host cells, effector immune cells also have a fairly strong negative relationship with healthy host cells.
We will analyze the system to ensure its chaoticity and complexity. This was done by performing the maximum Lyapunov exponent and sample entropy, and the results are presented in Table 3. The table shows that the average of maximum Lyapunov exponent value is positive, indicating a chaotic system. Additionally, the average sample entropy value for the three variables is less than 0.5, indicating the low complexity of the system. Therefore, it can be concluded that the system can be predicted accurately.

Linearization is key to accurately estimating the state using EKF. We achieve this by using Taylor expansion linearization up to the first order, from equations (6) - (8), which is as follows:

$$\dot{T} = (T_0 - T_0^2 - a_{12}T_0H_0 - a_{13}T_0E_0) + (1 - 2T_0 - a_{12}H_0)$$

**Table 3.** The Maximum Lyapunov Exponent and Sample Entropy for Itik Banks Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Maximum Lyapunov Exponent</th>
<th>Sample Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.05418</td>
<td>0.08273</td>
</tr>
<tr>
<td>H</td>
<td>0.04325</td>
<td>0.13738</td>
</tr>
<tr>
<td>E</td>
<td>0.06793</td>
<td>0.05774</td>
</tr>
<tr>
<td>Average</td>
<td>0.05512</td>
<td>0.09262</td>
</tr>
</tbody>
</table>
\[-a_{13} E_0 (T - T_0) - T_0 (H - H_0) - a_{13} T_0 (E - E_0),
\]
\[= (1 - 2 T_0 - a_{12} H_0 - a_{13} E_0) T - T_0 H - a_{13} T_0 E
\]
\[+ (T_0 - T_0^2 - a_{12} T_0 H_0 - a_{13} T_0 E_0) - (1 - 2 T_0 - a_{12} H_0
\]
\[-a_{13} E_0) T_0 + T_0 H + a_{13} T_0 E_0,
\]
\[\dot{H} = (r_2 H_0 (1 - H_0) - a_{21} H_0 (T - T_0)) + (r_2 - 2 r_2 H_0
\]
\[-a_{21} T_0) (H - H_0),
\]
\[= a_{21} H_0 T + (r_2 - 2 r_2 H_0 - a_{21} T_0) H + r_2 H_0 (1 - H_0)
\]
\[-a_{21} T_0 H_0 - a_{21} T_0 (T - T_0),
\]
\[E = \begin{pmatrix}
  \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} (T - T_0) + \frac{r_3 T_0 E_0}{T_0 + k_3} - a_{31} T_0 E_0
  \\
  -d_3 E_0
  \\
  \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} E_0 T + \frac{r_3 T_0 E_0}{T_0 + k_3} - a_{31} T_0 - d_3
  \\
  E
\end{pmatrix}
\]
\[= \begin{pmatrix}
  \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} E_0 T + \frac{r_3 T_0 E_0}{T_0 + k_3} - a_{31} T_0 - d_3
  \\
  \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} T_0 - d_3
  \\
  -a_{31} E_0 T_0 - \frac{r_3 T_0}{T_0 + k_3} - a_{31} T_0 - d_3
  \\
  E_0.
\end{pmatrix}
\]

The state-space representation from eq. (13) is

\[
x_t = \begin{pmatrix}
  \dot{H} \\
  \dot{E}
\end{pmatrix} = \Phi \begin{pmatrix}
  H \\
  E
\end{pmatrix} + B_t + w_t,
\]

\[
z_t = \begin{pmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 1
\end{pmatrix} x_t + v_t,
\]

where

\[
\Phi = \begin{pmatrix}
  \psi_1 & -T_0 & -a_{13} T_0 \\
  a_{21} H_0 & \psi_2 & 0 \\
  \psi_3 & 0 & \psi_4
\end{pmatrix},
\]

\[
B_t = \begin{pmatrix}
  \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} E_0 T
  \\
  \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} T_0 - d_3
  \\
  -a_{31} E_0 T
\end{pmatrix},
\]

\[
\psi_1 = (1 - 2 T_0 - a_{12} H_0 - a_{13} E_0),
\]

\[
\psi_2 = (r_2 - 2 r_2 H_0 - a_{21} T_0),
\]

\[
\psi_3 = \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} E_0 T,
\]

\[
\psi_4 = \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} T_0 - d_3,
\]

\[
\psi_5 = \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} T_0 - d_3,
\]

\[
\psi_6 = \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} T_0 - d_3,
\]

\[
\psi_7 = \frac{r_3 k_3 E_0}{(T_0 + k_3)^2} - a_{31} T_0 - d_3.
\]

In the numerical simulation, the Taylor series expansion of eq. (13) is carried out around \( T_0 = H_0 = E_0 = 0.01 \) and some parameters used are \( a_{12} = 1, a_{21} = 1.5, a_{13} = 2.5, a_{31} = 0.2, r_2 = 0.6, r_3 = 4.5, d_3 = 0.5, \) and \( k_3 = 1 \) [21]. The initial state and covariance are

\[
x_0 = \begin{pmatrix}
  0.01 \\
  0.01 \\
  0.01
\end{pmatrix} \quad \text{and} \quad P_0 = \begin{pmatrix}
  0.05 & 0 & 0 \\
  0 & 0.05 & 0 \\
  0 & 0 & 0.05
\end{pmatrix},
\]

respectively. The covariance of state error and measurement error is

\[
Q = \begin{pmatrix}
  0.05 & 0 & 0 \\
  0 & 0.07 & 0 \\
  0 & 0 & 0.8
\end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix}
  0.01 & 0 & 0 \\
  0 & 0.01 & 0 \\
  0 & 0 & 0.01
\end{pmatrix},
\]

respectively.

Figure 3 shows the EKF prediction results for the Itik Banks Tumor model/system. The figure consists of three subplots illustrating the dynamics of tumor cells (T), healthy host cells (H), and effector immune cells (E) over time. Each subplot compares the actual data against the prior and posterior predictions generated by the EKF. Figure 3a for tumor cells shows a strong agreement between the actual data and the prior and posterior predictions. The EKF captures the oscillatory nature of the tumor cell population well. The peaks and troughs of the predictions align closely with the actual data, indicating the EKF’s effectiveness in modeling the dynamic behavior of tumor cells. For healthy host cells, Figure 3b, the EKF predictions also demonstrate good alignment with the actual data. The periodic fluctuations in the host cell population are accurately mirrored by the prior and posterior predictions. This suggests that the EKF can reliably predict the interactions between healthy host cells and other components in the system. Figure 3c for effector immune cells indicates that the EKF predictions closely follow the actual data. The model successfully captures the oscillatory behavior and the sharp peaks characteristic of effector immune cell dynamics. The posterior predictions (blue dashed line) show a slightly better fit than the prior predictions (red solid line), demonstrating the EKF’s capability to refine predictions based on observed data. In the posterior step, the estimation results from the prior step are updated using actual data so that the estimation results of healthy host cells in the posterior step will approach their actual data. This correction is called data assimilation involving Kalman gain.

The estimation results of the rate of change of the three cells in the testing data are displayed in Figure 4. Figure 4a illustrates the EKF predictions for the tumor variable using testing data. The tumor data (T-actual) in green shows significant variations and peaks. The EKF prior (T-prior) and posterior (T-posterior) predictions, in red and blue, respectively, closely match the actual tumor data, demonstrating good predictive performance on unseen data. Figure 4b reveals the EKF predictions against testing data. While the EKF posterior (H-posterior) estimates (blue) capture the general trend of the actual data (green), there are some discrepancies, particularly at the troughs and peaks. The EKF prior (H-prior) predictions (red) show a larger deviation from the actual data, indicating the importance of posterior correction. Figure 4c shows the EKF predictions for the effector immune cells (E) with testing data. The actual data (E-actual) in green shows distinct peaks and troughs. The EKF posterior predictions (E-posterior) in blue track the actual data’s overall trend, though some oscillatory patterns are not captured accurately. The red EKF prior predictions (E-prior) exhibit more significant deviations, highlighting the improved accuracy of posterior estimates.

For the training data, the EKF posterior estimates for all three variables (Tumor, Healthy Host, and Effector Immune Cells) closely match the data during the training phase, indicating the
Figure 3. The EKF Prediction for Training Data: (a) Tumor, (b) Healthy Host, (c) Effector Immune Cells.

Figure 4. The EKF Prediction for Testing Data: (a) Tumor, (b) Healthy Host, (c) Effector Immune Cells.
filter’s effective adaptation to the known data patterns. Minor discrepancies are present but are generally small, suggesting a good fit to the training data. For the testing data, the EKF’s performance on testing data shows a strong general trend capture for all three variables, with posterior predictions often closely following the actual data. However, the deviations are more pronounced in the testing phase, especially for the Effector Immune Cells (E), indicating potential areas for model improvement or inherent unpredictability in the data. The posterior predictions (blue) are more accurate in the training and testing phases than the prior predictions (red). This difference emphasizes the importance of the update step in the EKF, which corrects the prior estimates using the actual observations. The oscillatory nature of the data is a significant challenge for the EKF, as seen in the Effector Immune Cells (E) variable. The EKF manages to capture the overall pattern but struggles with high-frequency oscillations. Despite some deviations, the EKF demonstrates robust performance in predicting the dynamics of the tumor, healthy host, and effector immune cells, making it a useful tool for such applications. In summary, the figures show that the EKF effectively tracks and predicts the behavior of key biological variables during both training and testing phases, with posterior estimates significantly improving prediction accuracy. The performance highlights the EKF’s potential for real-time monitoring and prediction in complex biological systems.

This study considers that data is available at each time step, so the prior estimation correction is performed when data is available. Therefore, the behavior of the three cells in the testing data resembles the estimation results in the training data for both the prior and posterior steps. The behavior of the estimated change rate in the three cells aligns with the simulation results conducted by [20]. However, when data is unavailable, the three cells are estimated using eq. (14). These estimations, illustrated in Figure 5, reveal certain insights into the prediction accuracy and areas that require improvement. Figure 5a, the EKF prediction deviates from the actual test data. This indicates that while the EKF can capture the general upward trend, it struggles with accurately predicting the intermediate fluctuations. Figure 5b, the prediction shows significant discrepancies, while the EKF prediction shows an upward spike. This suggests that the EKF is less effective in capturing sudden changes in the healthy host cell population, which may be influenced by external factors not accounted for in the model. Figure 5c, the EKF prediction shows large oscillations not present in the actual data, which remains relatively stable around zero. This indicates that the EKF struggles significantly with the effector immune cells, producing high-frequency oscillations that do not match the actual dynamics. The overall prediction error, measured by symmetric Mean Absolute Percentage Error (sMAPE), is 35.92%. This high error rate underscores the need for immediate correction once new data becomes available. The large deviations in the EKF predictions, particularly for the healthy host and effector immune cells, highlight areas for potential model improvement. The significant prediction error suggests that the EKF, in its current form, may not fully capture the system’s complex interactions and dynamics.

The Extended Kalman Filter (EKF) shows promising results in predicting the behavior of the tumor, healthy host, and effector immune cells in training and testing datasets. The posterior estimates produced by the EKF closely align with the actual
data during the training phase, indicating that the filter effectively adapts to known data patterns. The minor discrepancies observed are generally small, suggesting a good fit for the training data. In the testing phase, the EKF maintains its strong performance by capturing the general trend for all three variables. The posterior predictions (blue) typically follow the actual data closely, although deviations are more pronounced compared to the training phase, especially for the effector immune cells (E). This suggests potential areas for model improvement or highlights inherent unpredictability in the data, consistent with findings in the literature that underscore the complexity of tumor-immune interactions and the challenge of modeling them accurately [3, 16, 18].

The EKF’s ability to provide more accurate posterior predictions (red) compared to prior predictions in both the training and testing phases underscores the importance of the update step. This step corrects the prior estimates using actual observations, which is crucial for maintaining accuracy over time [20]. However, the oscillatory nature of the data, particularly for effector immune cells, poses a significant challenge. The EKF manages to capture the overall pattern but struggles with high-frequency oscillations. This aligns with previous studies that have highlighted the chaotic and oscillatory behavior of cancer and immune system dynamics, which can complicate prediction efforts [4, 14, 19]. Despite these challenges, the EKF demonstrates robust performance in predicting the dynamics of the tumor, healthy host, and effector immune cells. This makes it a valuable tool for real-time monitoring and prediction in complex biological systems, echoing similar applications of mathematical and computational models in cancer research [6, 11, 21].

The study also highlights that the EKF’s performance depends on data availability at each time step, allowing for prior estimation correction when data is available. This leads to similar behavior in the testing data observed in the training data for both prior and posterior steps. However, when data is unavailable, the estimation relies on pre-defined equations, which can introduce variability due to random noise. As shown in Figure 5, without the update step, the estimation of the three cells deviates more significantly from the actual data, leading to a larger prediction error. This necessitates prompt correction once new data becomes available to maintain prediction accuracy [19, 59].

5. Conclusion

This article successfully predicted the change rate of the tumor, healthy host, and effector immune cells using EKF. The state-space model representation is formed by linearizing the Itik-Banks tumor model. The EKF demonstrates strong performance in forecasting the behavior of tumor, healthy host, and effector immune cells within the Itik-Banks model. During the training phase, the EKF’s posterior estimates closely match the actual data, indicating effective adaptation to known data patterns with minor discrepancies. In the testing phase, the EKF continues to capture the general trend of all three variables, with the posterior predictions typically following the actual data closely. However, deviations are more pronounced in the testing phase, particularly for effector immune cells, suggesting areas for model improvement or highlighting inherent data unpredictability. The EKF’s ability to provide more accurate posterior predictions compared to prior predictions underscores the crucial role of the update step in correcting estimates using actual observations. Despite the challenges posed by the oscillatory nature of the data, especially for effector immune cells, the EKF demonstrates robust predictive performance, making it a valuable tool for real-time monitoring and prediction in complex biological systems. The necessity for prompt correction when data becomes available is emphasized, especially given the larger prediction errors observed when data is unavailable.

Future work should address the pronounced deviations observed during the testing phase, particularly for effector immune cells. This could involve enhancing the EKF’s capacity to handle high-frequency oscillations and chaotic behavior inherent in cancer dynamics. Integrating more sophisticated data assimilation techniques or hybrid modeling approaches may improve prediction accuracy. Expanding the model to incorporate more biological variables and external factors could provide a more comprehensive understanding of tumor-immune interactions. Investigating the impact of different noise models and refining the EKF parameters based on empirical data will also be crucial in enhancing the robustness and reliability of the prediction.

Author Contributions. Fitriyati, N.: A primary contributor to the drafting of the manuscript, linearization of the system, conducting numerical simulations for validation, and analysis of simulation results. Faizah, S. A.: Conduct linearization of the system and numerical simulations to validate and analyze simulation results. Sutanto, T. E.: Analysis and interpretation of simulation results, critical article revision for important intellectual content, and approved final manuscript.

Acknowledgement. The authors are grateful to the editor and reviewers for their careful reading, valuable comments, and helpful suggestions, which have helped them significantly improve this work’s excellence.

Funding. This research received no external funding.

Conflict of interest. The authors declare no conflicts of interest related to this article.

Data availability. The data supporting the findings of this study are available in the main references.

References


